Delft University of Technology

Engineering and Policy Analysis Master Thesis

Modelling a Race for Autonomy

A study of the system dynamics of the competition for autonomous military capabilities leading to potential arms races between two nations

Master Thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of
Master of Science
in Engineering and Policy Analysis
Faculty of Technology, Policy and Management

by

Laura Wilmes
(5234042)

To be defended in public on March 20 2023

Graduation Committee
Chairperson: Dr. Claudia Werker, Section Ethics and Philosophy of Technology
First Supervisor: Dr.ir. Willem Auping, Section Policy Analysis
External Supervisor: Rob van Waas MSc, TNO DSS Military Operations

March 4, 2023

TU Delft
Delft
University of
Technology

Executive Summary

With growing tensions in today's world, new powers rising and contesting old powers, two nations risk ending up in a so called security dilemma: One nation's increase in security threatens its opposing nation. The consequence could be the escalation of military buildups into arms races. With ongoing technological advancements, militaries around the world are expanding on autonomous capabilities that need less and less human control. While there are advantages of handing over dull, dirty, and dangerous tasks to robots, there are undesired consequences when nations engage in arms races for autonomous military capabilities: These include arming being a burden on economy and the piling up on capabilities with a varying degree of technological readiness.

In this research, I investigate under which conditions the developments of autonomous military capabilities between two fictitious nations may lead to arms races. These two nations reflect on the one hand a superior nation, with a strong economy and military, high ethical standards and a head start in the development of high quality technology. On the other hand, it's opponent is a nation that is economically and technologically inferior: it has a less strong but faster growing economy, a less strong military and less ethical standards which lacks behind in research and innovation. Further, strategies that these two nations can employ to influence arms races are investigated. These are strategies to prevent arms races such as bilateral disarmament agreements in the form of restricting the allowed level of autonomy of the two nation's military capabilities. Other strategies include targeting the opposing nation and its development of autonomous capabilities: This can be done in the form of limiting their access to funding, the components needed high quality technology and skilled researchers.

To answer the question under which conditions arms races arise and how to influence these, I construct a system dynamics model based on literature and expert consultations. With this model, I conduct experiments to generate a wide range of possible scenarios. These possible scenarios result from different initial conditions of two nations, such as different combinations of GDPs or different sizes of militaries. In all of these scenarios, I identify those scenarios that end in arms races which then allows to conclude on what causes these arms races.

The model results suggest that different types of arms races for autonomous military capabilities occur under specific conditions: the inferior nation lacks behind the superior nation and fills this initial gap not with a greater number of autonomous systems, but with higher quality autonomous systems, leading to an arms race in quality. When relating the model results to literature and more specifically to international relations theories, I find that arms races in the model happen under similar conditions as real world arms races.

When no nation employs any strategy to influence the arms races, the superior nation should avoid getting into one, since it loses more often then not. The contrary is true for the inferior nation, which wins the arms races most of the time. This suggests that the superior nation should try and spoil the race for its opponent.

Interestingly, bilateral disarmament in the form of regulations on the level of autonomy seem not to be the answer for the superior nation: some of the arms racing is reduced, while the superior nation now looses both racing and non racing scenarios. Plus, arms racing scenarios are not reduced significantly: This stems from the fact that the nations cannot compete in quality anymore and can only go for the quantity option. This then results in more quantitative arms races where the nations pile up a great number of autonomous systems. It remains debatable whether restricting the quality and quantity is in fact feasible, due to the dual use nature of autonomous robots and the uncountable nature of autonomy.

For a superior nation to prevent the occurrence of arms races, it is more effective, according to the model, to restrict an opponent in their technological development. Then,

even if the inferior nation strikes back, the superior nation is able to win in most scenarios; races and no races. Still, the inferior nation is able to win in the specific case, where the superior nation requires ”too much” human control for its autonomous military capabilities.

Generally, arms races can further be reduced when both nations avoid to be too ”reactive” to each other’s arming actions. This suggests, that a third party could step in and aim at relaxing the tensions between the two nations. Further, this third party could also aim at encouraging the inferior nation to raise its ethical standards. Then, the difference in ethical standards between the two nations is less big which gives the inferior nation less of a chance to use a kind of ethics strategy to win the arms race.

The model I built for this research is a quantified model based on large parameter ranges that aim at representing fictitious nations. I had to omit several aspects, such as active conflict or trading relations between the two nations. Future research includes to broaden the scope of the model and to include such elements. I then validate the model in this defined scope. Further, it could also be interesting to evaluate a regulation restricting the offensive capabilities of autonomous systems, as opposed to defensive capabilities. The current simulation model is not detailed enough to be able to test a regulation like this. For this, the autonomous systems and their specific applications would have to be modelled in more detail, with the technological developments that affect specifically these applications.

To conclude, the model can give insight into which decisions a nation or a third party can take with different goals in mind. These include, for the nations involved in a competition as a result from tensions, to either win the race towards autonomy or to prevent its opponent from winning it. A third party, interested in a peaceful world without having nations piling up autonomous military capabilities, might aim at preventing arms races whatsoever.

Acknowledgments

My sincere thanks go to my graduation committee, Dr. Claudia Werker and Dr.ir. Willem Auping, for enabling me to conduct this research. Further, I am grateful for all the smart people at TNO Military Operations and more specifically Rob van Waas and Patrick Steinmann for their support and for reminding me of the right questions to answer. Then, I would like to thank Prof.dr.ir. Mark Voskujil for his guidance in scoping this project. Finally, my gratitude goes to Lt Col Martijn Hädicke, without which my most interesting insights would not have been possible.

Laura Wilmes
The Hague, March 4, 2023

Contents

List of Figures

List of Tables

1 Introduction

The Cold War was characterized by arms racing. The two main powers of that time, namely the US and the Soviet Union, were amassing nuclear weapons as response to the other side ramping up on nuclear capabilities. This so called security dilemma refers to one country increasing their security (e.g., by increasing military capabilities) leading to less perceived security for another state (Mearsheimer, 2014; Kirshner, 2012). The number of weapons that one power builds up this way deters the other power to attack it, thus leading to an escalation of armed peace (Freedman, 2021). Each party knows that by attacking its opponent it would ensure its proper destruction. At the same time an arms race can potentially have a stabilizing effect as the nations focus on acquiring arms or technology and not on fighting (Gray, 1971).

After the Cold War, of which the US emerged as hegemon, new powers are rising that might threaten said hegemon's superiority. US hegemony is supported by economic strength, military might and cultural dominance. However, of other countries, China is challenging it in all aspects. (Mearsheimer, 2014) China, with its huge population, industry strength and its rising military strength is very likely to become a major power (Yuan, 2000).

According to international relations theory structural realism, a power rising potentially leads to new arms races. This results from the tensions with the current hegemon and the induced new security dilemma (Wohlforth, 2014). To consider a built up of capabilities as an arms race Caspary (1967) formulated three conditions, these being (1) perceived military insecurity, (2) the arming being a burden on economy and (3) perceived grievances.

The extent by which a security dilemma can be considered as arms race is not always clear. To illustrate the example of China and the US is given: The Taiwan Strait Crisis in 1996 made China perceive its military inferiority with regards to the US, when China was unable to deter the US from a major power demonstration in its own backyard (Teer et al., 2021). While both countries strive toward global technological leadership, impacts on economy can to some extent already be felt through export restrictions and tariffs on trade between each other (Fajgelbaum and Khandelwal, 2022). Grievances can be identified in Xi Jinping's speeches where he calls other nations to contest foreign dominance in Asia (China.org.cn, 2014) or in propaganda movies against Western organizations (chinascope, 2014).

Currently new military capabilities for a potential arms race are on the rise, namely robots, or in other words autonomous systems such as unmanned aerial systems (UAS) or unmanned ground vehicles (UGV). While the USA have been widely using UAS in the past, other countries are also getting on the train. It is estimated that in 2020, at least 102 countries, as well as some terrorist groups have active military drone programs (Gettinger, 2020; Voskuijl et al., 2020). In the six-week war in 2020 in the Nagorno-Karabakh region between Armenia and Azerbaijan, UAS have been used extensively to the extent that some scholars have called it the first ”Drone War” (Welt and Bowen, 2021). In the currently still ongoing war of aggression of Russia against the Ukraine (2023), ”kamikaze drones” are used in great numbers as loitering munitions against Ukraine as (Eslami, 2022).

The Dutch army is another good example of a military experimenting with autonomous systems. With their robotic and autonomous systems (RAS) unit, the Dutch army is working on concept development and experimentation with applications of autonomous systems (The Walking Soldier Podcast). Their systems are unmanned, while they stress the fact that "all weapons, including autonomous weapons, must remain under meaningful human control." (Michel Rademaker, 2022).

Ongoing research into artificial intelligence (AI) and machine learning will further increase the possible autonomy of the system and make human control less and less vital (Pallas Athena

Podcast). At the same time, research into battery technology, stealth or miniaturization contribute to the performance of these systems as military capabilities (Michel, 2020; Voskuijl, 2022).

These systems have a wide range of applications. Depending on the size and on the exact specifications of the different systems, they can be used for reconnaissance, target acquisition, swarming or as loitering munitions, to only name a few. (Michel, 2020). In these applications, due to their nature and depending on their level of autonomy, they have several advantages over manned systems, as they can take faster decisions and are less prone to fatigue.

The autonomy of these systems can be implemented to varying degrees. To measure the level of autonomy of a system, Sheridan and Verplank (1978) use a scale of 10 levels. On the lowest level on the scale, the human must take all decisions while the computer offers no assistance. On the highest level, the computer decides everything and ignored the human. A level in between, such as level 5 for instance, refers to the computer executing its own suggestions after human approval. Still, while technologically feasible, meaningful human control is an important element in autonomous military capabilities (Ekelhof, 2018). In addition, the ethical implications of using autonomous military systems can also affect their level of autonomy. Authoritarian states such as China might face less regulatory pressure in reducing the element of human control, while in the West, campaigns such as “Slaughterbots“ are already calling for a ban on autonomous military systems (Altmann and Sauer, 2017).

China is working on, what they call the ”intellectualization” of its military until 2035, with an emphasis on autonomous systems (Teer et al., 2021), while the US is employing its third offset strategy also with an emphasis on autonomous military systems (Fiott, 2016). Both countries are already frequently using UAS (Waldman, 2018; Wars, 2021) with varying degrees of autonomy.

The research and development into autonomy relies primarily on dual use technologies developed for civilian purposes (Lonardo, 2021). To illustrate, US based company Waymo and Chinese Baidu compete for the most autonomous self-driving car (McDonald, 2022). A reverse dynamic than during the Cold War is the result, during which commercial applications resulted from military innovations, such as GPS. AI for instance can be seen as an enabler for autonomous military systems similarly as the combustion engine for tanks (Horowitz, 2019). This leads to a ”spin-in” effect into military capabilities of civilian technology (Altmann and Sauer, 2017).

The effect of autonomous military systems on deterrence and on international stability is highly uncertain (Calcara et al., 2022). Some argue that autonomous military capabilities such as UAS could be the military revolution of the 21st (Altmann and Sauer, 2017; Horowitz, 2019; Sweijs and Osinga, 2021). Three combined factors underline this argumentation; (1) the fact that they are available at cheaper prices, (2) can blur the line between defensive and offensive purposes (which can increase the perceived military insecurity of a nation) (3) and have less life at stake. The latter condition could lower the threshold of actually engaging in a war, as the risk of losing lives is lower in the case of completely remote warfare (McKay et al., 2021). The first condition is given by the fact that a big part of innovation and development is conducted by the private sector for dual use technology. The second argument is debated by others arguing that UAS are not able to shift the offensive-defensive balance to the offensive side (Calcara et al., 2022).

Thus, it is imperative to investigate how a power competition between two states can escalate into an arms race for autonomous military systems. Arms races have been studied mostly in the context of nuclear weapons and the cold war. Plus, most of these models are of a qualitative nature as opposed to being quantitative models. With autonomous military capabilities being a relatively new phenomenon, there is more or less no research of this nature to be found.

This leads to the following research question: *How can the development of autonomous military capabilities of two states evolve and a potential arms race be avoided?*

This research question can be broken down into four parts:

1. Which considerations does a nation make to decide to develop its autonomous military capabilities?

2. How can the development of autonomous military capabilities of two states be modelled?

3. What is the role of technological developments such as autonomy and other factors in escalating or deescalating the development autonomous capabilities into a potential arms race?

4. Which strategies can both states take to influence the development of autonomous military capabilities and an escalating into a potential arms race?

This research problem links to my master program in Engineering and Policy Analysis at TU Delft. With this thesis, I study the interrelations between autonomous military technology and the societies which develop and invest into them. I then investigate the conditions under which arms races might occur by modelling a bipolar world. This world, in which arms races might take place represents a complex multi-actor system. Finally, to analyze the results, I need to make economic, political, and ethical reflections.

I conduct this thesis at the Dutch research institute TNO and more specifically its department for military operations in the context of strategic anticipation. The project team which I am working with is focusing on strategic defense analysis and on strategic foresight analysis. The objective of these analysis is to monitor and assess trends and their impact on security and the Dutch armed forces. Steps in doing so include trend analysis, driving forces identification and design of future worlds and scenarios. For this, they use many different scenario development methods, amongst others the qualitative modelling of a given problem statement. While qualitative methods suits themselves very well for gaining insight into the causal relations that underlie a problem, it cannot give a quantification of indicators. Therefore, with this thesis, I aim at exploring the role and potential added value of quantitative in strategic anticipation.

The report is organized as follows. First, section 2 identifies the methods to answering the introduced research questions. Then, section 4 presents the outcomes of applying these methods. Afterwards, section 5 provides an analysis of these results and finally, section 6 summarizes the findings of this study.

2 Methods

To investigate the research problem identified above, I first introduce System Dynamics (SD). For this, I define the scope of the model, as well as the dynamic hypothesis in the form of a qualitative causal loop diagram based on which further on the quantitative SD model will be built. I also introduce the methods to validate the model. I built the model with the input from experts, for which the approach of semi-structured interviews with professionals on the field of defense research; I use this input for the conceptualization and validation of the model. Furthermore, I introduce exploratory modelling and analysis, an approach that I use to identify those scenarios in which arms races happen. Finally, I explain the method of robust decision making to find strategies to avoid arms races.

2.1 System Dynamics

System Dynamics (SD) is particularly well suited to model systems with a high dynamic complexity, feedbacks, and delays (Sterman, 2002), as is the case for the system of this investigation. An SD model consists of interacting feedback loops, accumulations, flows and delays. This way, the causal relations that underlie the system are modelled. From these causal relations, the dynamic behavior arises endogenously (Kwakkel et al., 2013). In the system under investigation, these elements are all given. There are delays resulting from research and development, from procuring autonomous military systems and from responding to the opponent's actions. There are also accumulations due to the building up of military capabilities. Due to this, a stock-flow structure as often used in SD suits itself well for modelling the system under investigation. Finally, there is feedback, such as for instance between the arming up of one nation and the other nation.

While SD has been used in the past for modelling military capabilities, the aspect of autonomy has not been considered before. A first application of system dynamics for military capabilities was done by Forrester in 1984. The model was an implementation of Richardson's arms race equations that represents two states building up capabilities because of a desired superiority regarding the other state (Kreutzer, 1985). This model focuses on nuclear built ups between the US and the Soviet Union during the Cold War. The impact of autonomy on de-escalation or escalation mechanisms cannot be investigated using these models, as autonomous military capabilities are a relatively recent development.

2.1.1 Model Scope

As one cannot model the entire world, a scope needs to be set. The SD model to analyze the identified research problem reflects a bipolar world in which rising nation B contends hegemon country A. The focus is on these two fictitious nations and more specifically on their militaries. These two nations can be seen as placeholders for real nations. Nation A is the stronger nation. It has a stronger economy and as a result a higher GDP and therefore a higher defense spending. Nation B on the contrary, has a smaller GDP but therefore a faster GDP growth, as an analogy to B being a developing nation. In addition, nation A is the more efficient nation in terms of it military planning. It has faster acquisition cycles. Also, its capabilities are of higher qualities and as such have a longer lifetime before they need to be replaced. In addition, nation A has higher ethical standards as nation B. Hence, it needs a greater number of personnel for its autonomous systems than nation B and it also trains its personnel longer.

The tensions between hegemon nation A and rising nation B lead to a security dilemma (Wohlforth, 2014) and can result in an arms race for high tech military capabilities (Caspary, 1967). More specifically, the militaries of nation A and nation B are ramping up on autonomous military capabilities of which the development is also driven by dual use technology. The causal loop diagram can be seen in Figure 1. Due to nation A being the stronger nation, it has a head start in the race for autonomy.

The system boundaries determine what is modelled and what is omitted. I deliberately omit and hence do not include in the model scope the cultural differences between nation A and nation B, national internal stability, other disruptive technology, or other military engagements. I also do not include an outright war or third nations that could figure as strategic partners in trade or defense. Further, I omit diplomatic channels and other ways of conducting foreign policy between nation A and B and do not model trading relations between nations A and B.

Originating from outside of the system is economy. Therefore, in the world of the model, the GDP is exogenous, and the economy is simply assumed to continue growing more or less fast. The same goes for scientific research; I assume the level of technological sophistication and of AI to increase. This is assumption is valid, as not only is the development of AI and autonomy driven by military demand, but also by civilian demand.

There are three endogenous subsystems, meaning that I model in detail and of which the behavior is generated by the model structure: the race, technology, and planning submodels that interact with each other. These three subsystems capture the elements a nation needs in order to build autonomous military capabilities, these being the needed investments and funding, the skilled personnel for developing autonomy and the needed critical components to build autonomous systems (MoD, Concepts and Doctrine Centre, 2018). The race submodel represents the security dilemma. The technology submodel comprises the development of high quality autonomous military systems. This is on one hand based on artificial intelligence and the availability of highly skilled scientists that can enable this research. On the other hand, it depends on the access to critical components that allow for greater technological sophistication of the autonomous systems, such as semiconductors and sensors. The planning submodel consists in determining defense spending due to the security dilemma. Further included in defense spending are costs related to personnel and other military capabilities.

2.1.2 Dynamic Hypothesis

Before building the quantitative SD model, a dynamic hypothesis is needed. This dynamic hypothesis reflects the conceptualization of the system and links the assumed structure of the system to its assumed behavior, all within the scope of the model as defined above.

The underlying assumption is the escalation archetype. The escalation archetype describes two balancing loops: Due to an increase of an actor A's performance relative to an actor B, the threat posed to an actor A decreases. At the same time, this increases the threat posed to actor B such that this actor takes actions that decrease the advantage in relative performance of A compared to actor B. This then results in an increase in threat to actor A which in turn takes actions to again increase its advantage. This illustrates that if decoupled, both actors would in fact find an equilibrium. Due to the interaction of these two loops however, the result is a vicious circle of potential escalation and exponential growth of the behavior that is aimed at decreasing the perceived threat. The escalation archetype is hence naturally suitable for describing the security dilemma (Mearsheimer, 2014; Kirshner, 2012). By increasing its security, a nation decreases the security of its opponent. The opponent in turn reacts by increasing its own security which then decreases the security of the first nation.

The causal loop diagram (CLD) as can be seen in Figure 1 is an extension of the escalation archetype and depicts several feedback loops where the causal relations of nation A are depicted in red and those of nation B in blue. This causal loop diagram is derived in an iterative process from reviewing literature and conversations with subject matter experts. It represents the dynamic hypothesis in form of a qualitative model based on which in the next modelling stage, the SD model as a quantitative model in its stock-flow structure will be built.

In the context of an arms race for autonomous military capabilities, I understand ”increasing one’s own security” as increasing the quantity and quality of one’s current autonomous military capabilities. I assume that by perceiving an opponent’s stock of existing autonomous military capabilities and the amount of autonomous military capabilities under development, a nation wants to invest into increasing its own stock of autonomous military capabilities.

Due to the security dilemma, one nation ramping up on autonomous military capabilities will almost always escalate into an arms race (reinforcing loop R2). One nation increasing its own security decreases the other nation’s security, while each nation bases itself on the perception of the other nation’s power. The exactness of this perception has a big influence on a potential escalation. If A only *thinks* that B is more powerful than it actually is, the building up on capabilities escalates without there actually being a security dilemma to begin with. Hence, building trust between the nations can help reduce the threat by a nation and avoid an unintended escalation. To illustrate, by mutually agreeing on a certain number of capabilities and a certain level of autonomy, for instance through arms control agreements, an equilibrium or de-escalation might be achieved.

With more R&D and more technological sophistication, the security dilemma can be further enforced. An increased level of autonomy of the developed autonomous military capabilities will further increase the threat posed by a nation (R1 for nation A and R3 for nation B). This is hard to balance since autonomous military capabilities pose an additional challenge to trust. It is difficult to quantify a given level of autonomy of a system and the nations will not share their software with each other. Also, the doctrine with which a nation plans to employ its autonomous systems will most probably be classified. Therefore, I assume an increase in autonomy to reduce the potential balancing effect of trust as deescalating mechanism.

The speed of the escalation depends on the delayed spin-in of civilian technology into military capabilities. Since technological development is driven by the security dilemma as well as by civilian demand, there will be spin-in of civilian technology into military capabilities. In the case where A and B share and trade their civilian dual use technologies, nation A will get an advantage if their spin-in is shorter than nation B’s, for instance when nation B imposes regulations to restrict the level of autonomy too early. This results in nation A beating nation B with its own technology. In addition, the rising and less ethical nation B could have an advantage in developing autonomy resulting from the massive amount of data it collects from surveillance of its population. If, however, nation A’s spin-in is shorter than nation B’s, nation A does not bother about sharing its technology with B: By the time that nation B has managed to spin-in nation B’s technology into their military capabilities, nation A has already developed more advanced technology.

Trade can have both a deescalating and an escalating effect. Engaging in trading relations between the nations decreases the threat posed by the nations. While initially I included trade in the dynamic hypothesis, during the conceptualization of the model I decided not to, as the focus is on military autonomous capabilities which the two nations will not be trading with each other (see Appendix B).

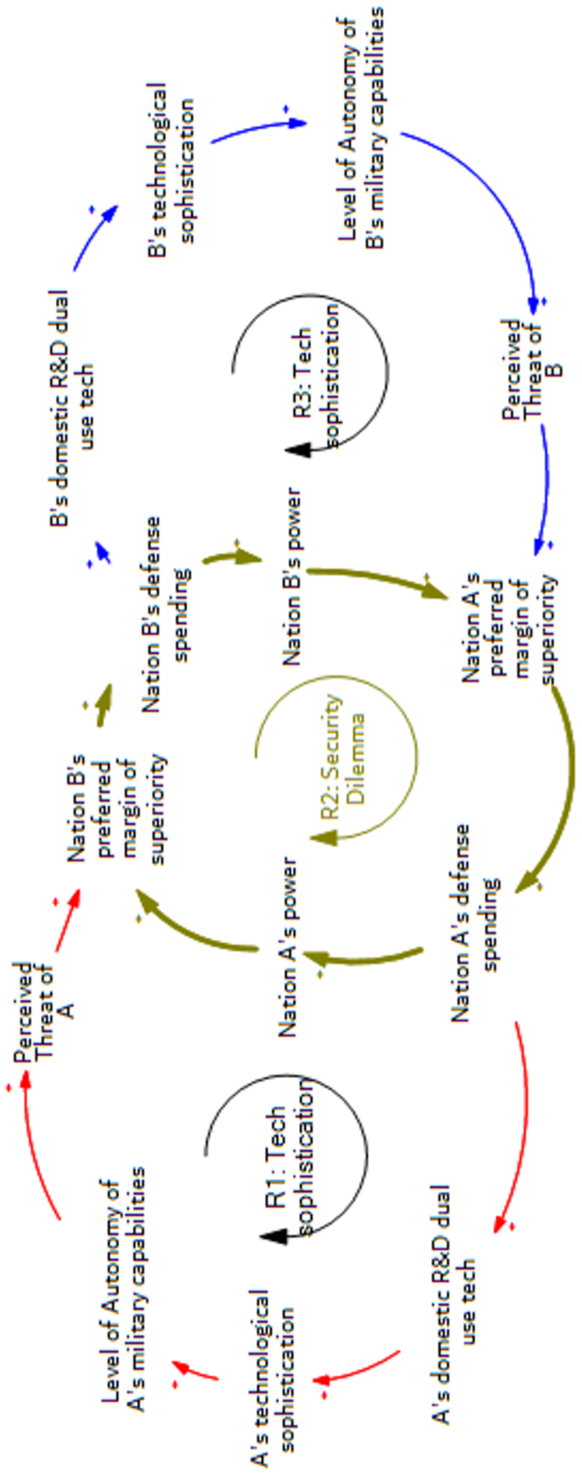

A's domestic R&D dual use tech
A's technological sophistication
Level of Autonomy of A's military capabilities
R1: Tech sophistication
Perceived Threat of A
Nation A's defense spending
Nation A's power
Nation B's preferred margin of superiority
R2: Security Dilemma
Nation A's preferred margin of superiority
Nation B's power
Nation B's defense spending
B's domestic R&D dual use tech
R3: Tech sophistication
Perceived Threat of B
B's technological sophistication
Level of Autonomy of B's military capabilities
+
+
+
+
+
+
+
+
+
+
+
+
+
+
+
+

Figure 1: Causal Loop Diagram depicting the dynamic hypothesis based on which the SD model will be built.

Based on the dynamic hypothesis as represented by the qualitative causal loop diagram, I construct the quantitative SD model using stock-flow structures within the model scope. Thus, I can quantify the qualitative derived parameter relations as depicted in the causal loop diagram and use the SD model as a simulation model. Thus, with the quantification, I can also analyze behavior over time and how the loops, delays and accumulations interact.

2.1.3 Model Validation

A constructed model, qualitative and quantitative, needs to be validated to determine to what extent its results are useful. For this the validity of the model depends on the purpose and hence it needs to be determined whether the model is fit for purpose: I should be able to use it to answer the research question, namely how the development of autonomous military capabilities of two nations can evolve and potential arms racing behavior be influenced (Sterman, 2000). With this section, I expand on the validation process of the CLD as presented above and of the SD model that I will present in section 3.

I conducted structure verification of the qualitative CLD model with experts. For this, I did interviews and workshops with the consulted interviewees to discuss the results of the model structure (Senge and Forrester, 1980). This results in an iterative modelling process were based on the feedback obtained during the workshops, the model was adapted until deemed fit for purpose. With a CLD that is fit for purpose, I could construct the quantitative SD model in section 3.

For validation of the SD model, the distinction is made between structure verification tests on the one hand and model behavior verification tests on the other hand Senge and Forrester (1980). With structure verification tests, I aimed at verifying whether the model structure and the model parameter reflect what can be observed in the real world. I conducted these tests based on the available literature as well as during reviews with expert interviews, resulting in an iterative modelling process. I also checked boundary-adequacy with the aim at determining whether the structural relationships are necessary for answering the research question of this report. To do so, I frequently revised the dynamic hypothesis, resulting in an iterative modelling process. I also checked Dimensional consistency on a repetitive basis. The results of these tests can be found in section 3.

Model behavior tests aim at analyzing the possible behaviors that the model can generate (Senge and Forrester, 1980). After conducting open exploration to see which behaviors the model can generate, I conducted behavior-sensitivity tests to determine how the model behavior depends on certain combinations of parameter values. For doing so, I determined a plausible range for each parameter as well as different lookup variations and I simulate the model to generate different outcomes over the entire range of different parameter combinations. Then, I analyze the resulting defense spendings and check to what extent these reflect real world behavior, where I consider any defense spending of greater than 46% of GDP as extremely high (Collier and Hoeffler, 2002). I conducted the model tests with the python module EMA workbench (Kwakkel, 2022). The results of these tests can be found in section 3. A model's structure should also permit extreme conditions and the model's results reflect reality, according to Senge and Forrester (1980). Hence, I also conducted extreme condition tests that I report on in Appendix C.

2.2 Semi-structured Interviews

Interviews can provide necessary data when the phenomenon of interest is infrequent and alternative data sources are scarce. As arms races are rather infrequent, this description holds through for investigating the research question of this study. Therefore, I conducted semi-structured interviews with experts (Eisenhardt and Graebner, 2007; Flick, 2022). I used the data generated during the interviews for model specification and for model validation. I consulted different groups of experts for these different steps with the aim of an unbiased validation process.

Involving experts into research can enhance the quality and validity of the produced results. It allows to investigate the system from within and incorporate the experiences, perspectives, and arguments of the stakeholders (Király and Miskolczi, 2019). In addition, it allows to incorporate the experts' "mental models" into the structure of the SD model (Sterman, 2000). In addition, by consulting different experts, different "mental models" can be considered and integrated into one.

To limit bias in the data collection, I identified experts with different background during the exploration of the field of research (Eisenhardt and Graebner, 2007). Resources consist in conference articles, journals, reports and newspaper articles, podcasts, and symposiums, the latter two mostly from Dutch sources. Due to the strict time frame of this thesis project, I could not reach out to experts for instance from other cultural and national backgrounds and some experts were not available for an interview in the short amount of time.

For the interviews, the experts I consult are all based in The Netherlands and all affiliated to the defense and security domain. They are mostly working at the Dutch research institute TNO in its unit Defense, Safety and Security (DSS) at which I conducted this thesis. TNO is the Netherlands Organization for Applied Scientific Research and its DSS unit is working on projects that concern defense and national security. Hence, the people working here generally have a background in the defense and security domain. Therefore, they, their knowledge and their advice constituted a good starting point for me to conceptualize the model. First, I interviewed defense modelers who, based on their experience with modeling different defense related topics, could give me insight into the potential structures of an arms race model. They explained which archetypes are to be expected and which dynamics could be expected. I also interviewed the strategic defense analysts which could give me advise on the conceptualization and strategic considerations that nation states take when deciding on capability planning. Those interviews gave insight into theories from international relations and identified which aspects are crucial to consider when modeling an arms race.

With these experts, I conducted semi-structured interviews. These are based on the use of open-ended questions to get the experts to expand on their answers and perspectives that are of interest for me. The focus is on getting a better understanding information that has been obtained after reviewing literature and other sources (Flick, 2022). In preparation to the interviews, I consulted resources, and I identified the questions that I could not answer. A list of questions can be found in Appendix A. The questions are categorized into those regarding the specifics of technological developments of autonomous military capabilities, questions about the geopolitical implications of autonomous military capabilities and finally those about how specific systems such as UAS as military capabilities are applied. However, it needs to be noted that due to the semi-structured nature of the interviews, I also discussed some topics freely with some of my experts, without sticking to a list of predefined questions. Some of my experts also agreed on several consultations which enabled me to reiterate on several questions and get a deeper understanding on the research topic.

Throughout the model building process, I use the generated data in an iterative way (see Figure 2). This also made it possible for me to make use of their knowledge during the specification of the model, the validation phase as well as for the interpretation of the results (Senge and Forrester, 1980; Charmaz, 2014).

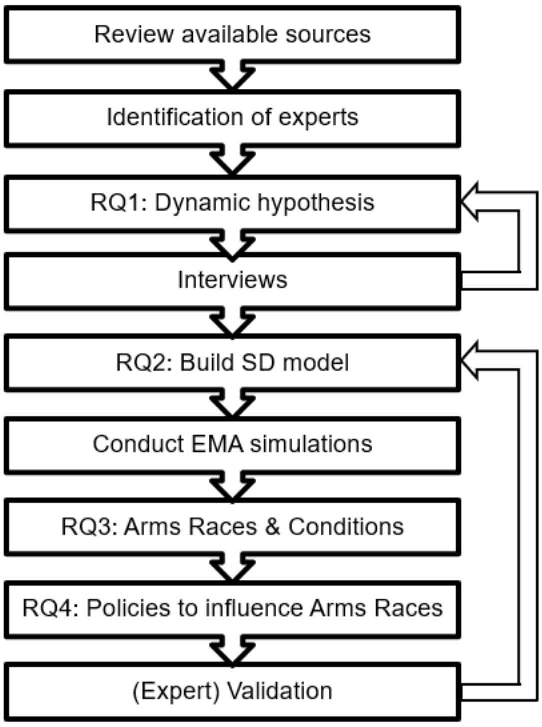


Figure 2: Research Flow showing the iterative process of model building. (Different) experts are consulted twice in the process; in the first cycle for conceptualizing the model and in the second cycle for validating the model.

2.3 Scenario Discovery

To analyze the role of technological and other developments and factors on a potential escalation, I use exploratory modeling and analysis (EMA). EMA helps to overcome the inherent uncertainty of model parameters and structure and understand the behavior of the system. EMA relates to investigating how uncertainties in a model and/or policy levers map to outcomes of the model and is used to conduct computational experiments (Walker et al., 2013). Here, uncertainties are for instance the level of technological sophistication and levers include the level of defense spending.

In particular, scenario discovery will enable to understand under which circumstances (i.e., combination of uncertainties and levers) the arms race escalates. For this, I use the Patient Rule Induction Method (PRIM) algorithm to identify future worlds of interest (Kwakkel, 2022). These are futures in which the arms race escalates and both states pile up on autonomous military capabilities.

I define arms races as scenarios where the defense spending of a nation increases by more than 50% over a period of 5 years (Diehl, 1985). Both bilateral as well as unilateral buildups are of interest (Boswinkel and Sweijs, 2022). In the former, both sides are piling up autonomous systems while in the latter case, it is only one of both nations that is increasing its defense spending significantly. There are two different types of arms races: qualitative and quantitative arms races (Intriligator and Brito, 1984). In a quantitative arms race, nations pile up large amounts of systems. They are measured by the stock of systems a nation acquires over time. In a qualitative arms race, nations aim at acquiring systems that are highly autonomous. These are measured by looking at the effectiveness of these systems, stemming from their technological sophistication and their level of autonomy.

Also of interest are the outcomes of the arms races, namely which nation wins the competition for autonomous military capabilities. As quantity as well as quality are important indicators, I compute a nation's weighted quantity: For this, at the end of the simulation time, I multiply the number of systems a nation has accumulated (and hence weighted) with their overall effectiveness. The winning nation in each scenario is thus the one with the highest weighted quantity in capabilities. Scenarios of interest are then those where hegemon nation A wins. Here, I choose the point of view of nation A as I interpret this nation as the ”Western” nation. As I construct this model with my ”Western” assumptions and bias, this seems to be natural point of view to take.

PRIM clusters the computational experiments conducted with the SD model. First, the uncertain parameter ranges are determined and then, simulation runs are conducted on the SD model using all possible parameter combinations over these uncertainty ranges (e.g., ratio of technological sophistication ranging from 0.5 to 5 at the same time where desired margin of superiority ranges from 1 to 5). After this, PRIM is used on these runs. PRIM generates boxes in regions of these parameter combinations where a particular value of interest is particularly high or low (in this case e.g., the stock of autonomous military capabilities and technological sophistication).

Thus, I can identify factors that escalate or deescalate the built up for autonomous military capabilities. For each box that is generated by the PRIM algorithm, the rules that define the box can be examined. It can be inferred which parameter range and parameter combination cause this run to happen and form the particular discovered scenario of interest. From this, I can finally conclude under which technological advancements and other factors in the model an escalation of the built up of autonomous military capabilities into an arms race takes place and in which it does not (Groves and Lempert, 2007).

When PRIM does not lead to conclusive results, I can still manually analyze the parameter ranges and combinations that lead to arms races. For this, I separate scenarios with arms races and without arms races. For each parameter in the uncertainty space, I take the average once for arms racing scenarios and once for scenarios without arms races. Then, I take the relative difference between both cases. If there is a significant difference of an average parameter value in the arms racing scenarios versus no arms racing scenarios, I determine this parameter to be characteristic of an arms race.

2.4 Robust Decision Making

To influence the occurrence of arms races, nations employ strategies to reduce unwanted outcomes. These strategies include decisions on funding, arms control agreements such as regulations on the level of autonomy of their systems, and sanctions. The decision for adopting a specific strategy depends on the assumptions of the model. Due to its inherent uncertainties, the effect of a policy on the system is also uncertain (Groves and Lempert, 2007). Consequently, the decisions taken need to be robust in the light of these uncertainties.

With Robust Decision Making (RDM), strategies are chosen that perform well over a wide range of future worlds. The chosen strategy is thus one with a small regret compared to other available strategies. Regret is the difference between performance of the chosen strategy and the performance of other available strategies (Lempert et al., 2006).

There are two different types of strategies in the model: The first type are the bilateral regulation strategies where each nation restrains itself from engaging in arms races. Here, the nations place a cap on the level of autonomy, at different three different time steps during the simulation time. Thus, there is (1) the static regulation where the level of autonomy is restricted from the start on, (2) the midway strategy where the level of autonomy is restricted halfway through the simulation time and (3) the adapting strategy where the level of autonomy is restricted as soon as one nation exceeds a given level of autonomy.

The second type are the spoiling strategies where each nation targets its opposing nation's development of autonomous capabilities. I implement three different spoiling strategies, these being (1) the funding strategy aimed at quantitative arms races in which nations race to acquire huge amounts of capabilities, (2) the quality for addressing the qualitative arms races in which nations race to acquire highly autonomous systems and (3) the combined strategy.

For the first step for RDM, I conduct computational experiments with the SD model where the outcomes of all possible strategies are computed over the possible parameter ranges (e.g., investing 0.5% of GDP into defense spending up to investing 20%). At the same time, I compute all possible future worlds over all possible ranges (e.g., ratio of technological sophistication ranging from 0.5 to 5).

To choose an initial possible strategy, I rank these, and I choose the one performing the best over a wide range of possible futures. The best performing strategy is the one that results in the least number of arms races and where hegemon nation A wins most the competition for autonomous military capabilities with rising nation B most often.

In the second step, I identify the vulnerabilities of this strategy using scenario discovery as described above to find the clusters in the computational experiments where the chosen strategy does not perform well. To give an example, the implementation of sanctions on the opposing nation might generally help to reduce arms race. However, it might not in those scenarios, where the other nation spends more than 10% of its GDP.

In the third and final step, I take other strategies into consideration that could make up for the vulnerabilities of the chosen strategy and thus perform well in these worlds where the chosen strategy does not. To illustrate, in these scenarios, arms control agreements between the two nations could perform better and prevent the built up of autonomous military capabilities into an arms race. (Lempert et al., 2006)

3 Model Specification

In this section I address the first steps in answering the research question of this thesis; namely how the development of autonomous military capabilities of two states can be modelled, based on the considerations nations make when decide to develop autonomous military capabilities. I built the quantified SD model based on the qualitative causal loop representing the dynamic hypothesis from section 2. The SD model consists of three different submodels with different stock flow structures, these being the arms race, the technology, and the planning submodel. I used Vensim for the modelling process, with, after some trial and error, Euler as integration method and a time step of 0.00390625 years with a simulation time of 30 years. I introduce the three different submodels separately in the next sections. I will only show the submodels for nation A, while nation B has the same structure as nation A but with different parameter values. The model itself with the documentation of equations can be found on my repository on GitHub.

3.1 Arms Race submodel

The first submodel, the racing submodel, represents the security dilemma based on a first implementation of Kreutzer (1985) by Willem Auping. The underlying assumption of this structure is the escalation archetype in the form of the dynamic hypothesis from section 2.

The racing submodel, as all other submodels as well, is mirrored. In Figure 3, I explain the submodel for nation A and thus, only the aging chain of autonomous military capabilities in the form of a stock flow structure for hegemon nation A can be seen. For rising nation B, it looks the same. Only the parameters are different to make a difference between nation A being the superior nation and B being the rising nation.

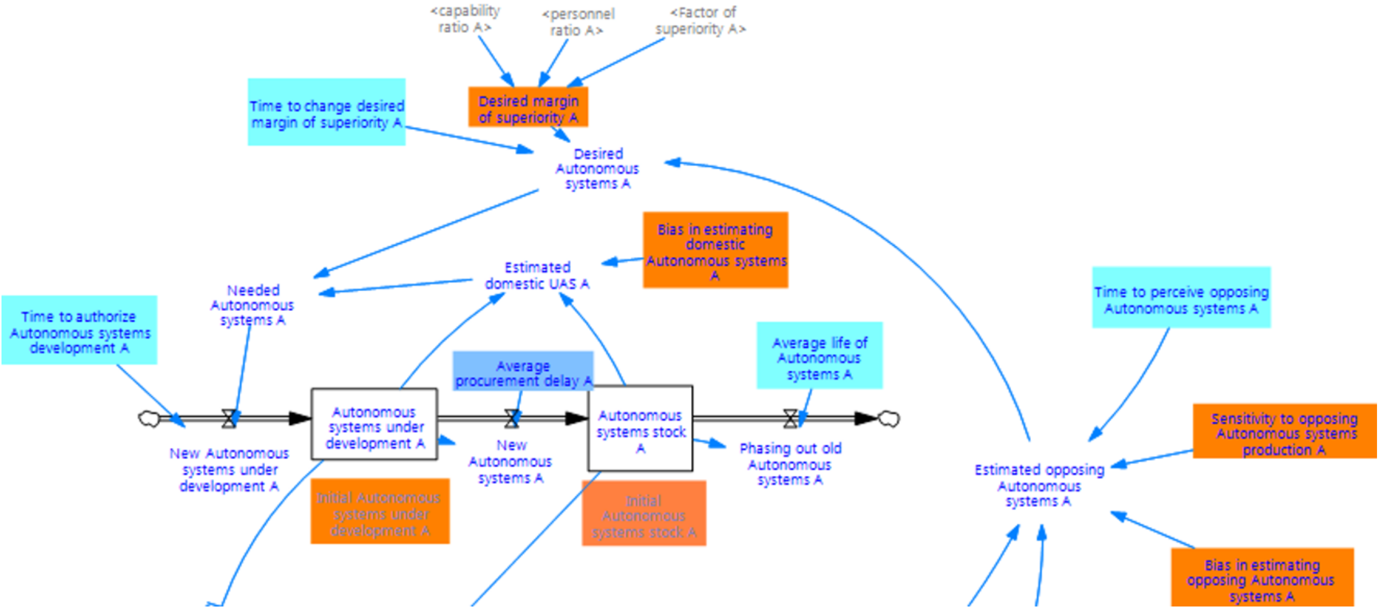

<capability ratio A>
<personnel ratio A>
<Factor of superiority A>
Time to change desired margin of superiority A
Desired margin of superiority A
Desired Autonomous systems A
Bias in estimating domestic Autonomous systems A
Estimated domestic UAS A
Needed Autonomous systems A
Time to authorize Autonomous systems development A
Time to perceive opposing Autonomous systems A
Average procurement delay A
Average life of Autonomous systems A
Autonomous systems under development A
Autonomous systems stock A
New Autonomous systems under development A
New Autonomous systems A
Phasing out old Autonomous systems A
Initial Autonomous systems under development A
Initial Autonomous systems stock A
Estimated opposing Autonomous systems A
Sensitivity to opposing Autonomous systems production A
Bias in estimating opposing Autonomous systems A

Figure 3: The racing submodel depicting the security dilemma from the side of nation A. Nation B is modelled with the same structure.

The quantity of autonomous military capabilities of a nation is modelled using an aging chain with two stocks. The first flow of this aging chain is the ”New autonomous military capabilities under development” which refers to the amount of new autonomous military capabilities that a nation needs after a given time that it takes the nation to authorize the new development of autonomous military capabilities. After this time delay, autonomous military capabilities flow into the stock of ”autonomous military capabilities under development” where they are accumulated. After the time delay that it takes to file all the necessary bureaucratic procedures to actually procure new autonomous military capabilities, these flow into the ”autonomous military capabilities stock” where they accumulated. After a given time, their lifetime is over and phase out.

Autonomous military systems such as small UAS have the potential of bringing quantity back, as they are comparably cheaper as other military capabilities (Pallas Athena Podcast). Thus, the amount of needed autonomous military capabilities that will be developed is determined by comparing the estimated amount of autonomous military capabilities of a nation and its desired amount of autonomous military capabilities. The existing quantity of autonomous military capabilities that a nation has (from the stocks of autonomous military capabilities under development and the autonomous military capabilities stock) is corrected with a bias factor to make up for estimation errors about the nation’s opponent. The desired amount of autonomous military capabilities depends on what the nation thinks the opposing nation has as autonomous military capabilities (Kreutzer, 1985). Perception factors are introduced to make up for estimation errors of the other side. For estimating the opposing time, a given delay time is also applied, due to bureaucratic and political procedures.

The amount of desired autonomous military capabilities is calculated by applying a desired margin of superiority (Kreutzer, 1985). This refers to by how much a nation wants to be superior to the other nation and is calculated based on the factor of superiority resulting from the technological superiority and based on the ratios of personnel and other military capabilities. This way, a nation’s military power is represented not only by their autonomous military capabilities but also their strength in personnel and other capabilities (In.Europa., 2017). The factor of superiority, referring to the technological superiority is calculated in the technology submodel. It represents a nation’s desire to be superior not only in their quantity of autonomous military capabilities but also in their quality of the autonomous military capabilities (Intriligator and Brito, 1984). Quality of autonomous military capabilities is understood as their technological sophistication as well as their level of autonomy.

The need for new autonomous military capabilities is covered either by available funds that are allocated to the development of autonomous military capabilities. If, with an ongoing arms race, these available funds are not enough due to an increase in the ”autonomy gap” between estimated domestic and estimated opposing autonomous military capabilities, the available defense budget needs to be increased. This is calculated in the Capability Planning submodel.

To influence arms races, policy switches for the ”funding strategy” and part of the ”combined strategy” are implemented in this submodel. These are employed by the two nations to target their opposing nation to spoil the race for autonomy. These measures aim at quantitative arms races (Intriligator and Brito, 1984). The nations will be able to restrict their opponent’s funding for autonomous military capabilities and as such the quantity of systems they can acquire. The policies are activated depending on which option the opposing nation chooses for its race for autonomy. If the opponent goes for the qualitative option, funding is not necessarily critical to its success. If, however, the opponent chooses the quantitative option, funding becomes critical to it (MoD, Concepts and Doctrine Centre, 2018). Thus, as soon as the quantity of autonomous capabilities of the opposing nation surpasses the domestic capabilities, the policies are activated. Then, without the needed funding, a nation cannot acquire new systems.

3.2 Technology submodel

Autonomous military capabilities as a capability combine quantity of the platforms with quality stemming from the underlying technology. Artificial intelligence and the related autonomy are amongst the top three technological developments that increase the effectiveness of autonomous military capabilities (MoD, Concepts and Doctrine Centre, 2018).

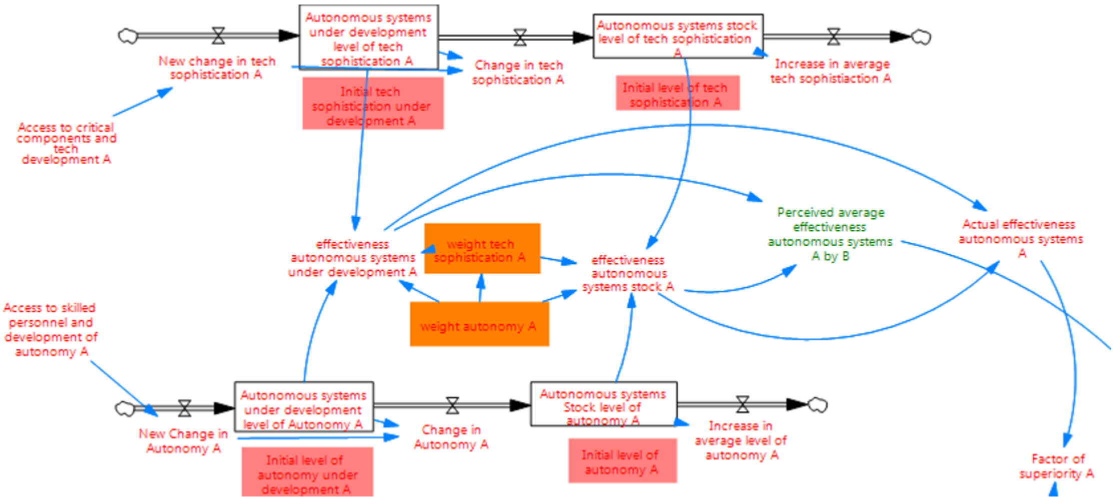

Autonomous systems under development level of tech sophistication A
Autonomous systems stock level of tech sophistication A
New change in tech sophistication A
Change in tech sophistication A
Increase in average tech sophistiaction A
Initial tech sophistication under development A
Initial level of tech sophistication A
Access to critical components and tech development A
Perceived average effectiveness autonomous systems A by B
Actual effectiveness autonomous systems A
effectiveness autonomous systems under development A
weight tech sophistication A
effectiveness autonomous systems stock A
weight autonomy A
Access to skilled personnel and development of autonomy A
Autonomous systems under development level of Autonomy A
Autonomous systems Stock level of autonomy A
New Change in Autonomy A
Change in Autonomy A
Increase in average level of autonomy A
Initial level of autonomy under development A
Initial level of autonomy A
Factor of superiority A

Figure 4: Technology Submodel showing the co-flows of the development of autonomy and technological sophistication of the autonomous military capabilities of nation. Again, the structure for nation B is the same.

The factor of superiority hence reflects by how much a nation needs to be superior due to a higher level of effectiveness of the opposing nation's autonomous military capabilities. This effectiveness is calculated based on the perceived average level of technological sophistication and autonomy of the opponent's autonomous military capabilities. This is then compared to the average level of technological sophistication and autonomy of the nation's own capabilities (see Figure 4).

To calculate the overall effectiveness, weighting factors reflect how much importance each nation gives to either technological sophistical or autonomy. These two characteristics of a nation's autonomous military capabilities are modelled using a co-flow structure based on (Sterman, 2010) (see Figure 4). These two co-flows behave in the same way and with the same time delays than the main stock flow structure in the race submodel.

While quantity is achievable without necessarily having access to highly skilled scientists and researchers or to critical components, the quality of the systems is very much dependent on these elements (MoD, Concepts and Doctrine Centre, 2018). Thus, the achievement of general technological sophistication depends on the availability of critical components that are needed for high quality autonomous military capabilities. These are for instance semiconductors or sensors. The achievable level of autonomy depends on the development of artificial intelligence. For this, a nation also needs to have access to highly skilled personnel that can conduct this research. As was mentioned in section 2, these two elements are exogenous for the dual use nature of autonomous robots. Not only is autonomy developed for military purposes but also for civilian demand.

Several policy switches are implemented in this submodel, focusing on qualitative arms race (Intriligator and Brito, 1984). When the two nations restrain themselves from engaging in an arms race, the policy switch for the "regulation strategy" is switched on. This enables both nations to bilaterally put a given cap on the autonomy of their military capabilities under development. This cap of a level of autonomy of 5 when using the scale of Sheridan and Verplank (1978) refers to the system being able to make and execute its own suggestions after human approval. This can be seen as “meaningful human control” (Ekelhof, 2018). This cap on autonomy can be placed at different time steps during the simulation time: from the beginning on, halfway through the simulation, or as soon as one nation exceeds level 5.

Otherwise, when the nations aim at winning the arms race, they target their opponent and impede it in its efforts in the race for autonomy. This reflect the ”quality strategy” and part of the ”combined strategy” and restricts the opposing nation from access to high quality systems. Hence, there are two switches that restrict the opposing nation from access to critical components on the one hand and to skilled personnel on the other (MoD, Concepts and Doctrine Centre, 2018). These are activated when the opponent goes for the quality strategy as opposed to the quantity option for winning the race for autonomy. Then skilled personnel and access to critical components become critical for its success. Thus, these policies are activated when the level of effectiveness of the opposing nation’s autonomous capabilities surpasses the domestic one.

3.3 Capability Planning submodel

The Capability Planning submodel reflects how much budget a nation allocates to defense spending. Some of the structure is based on a prior model of Willem Auping. Of this total defense budget, a fraction is spent on autonomous capabilities (see Figure 5).

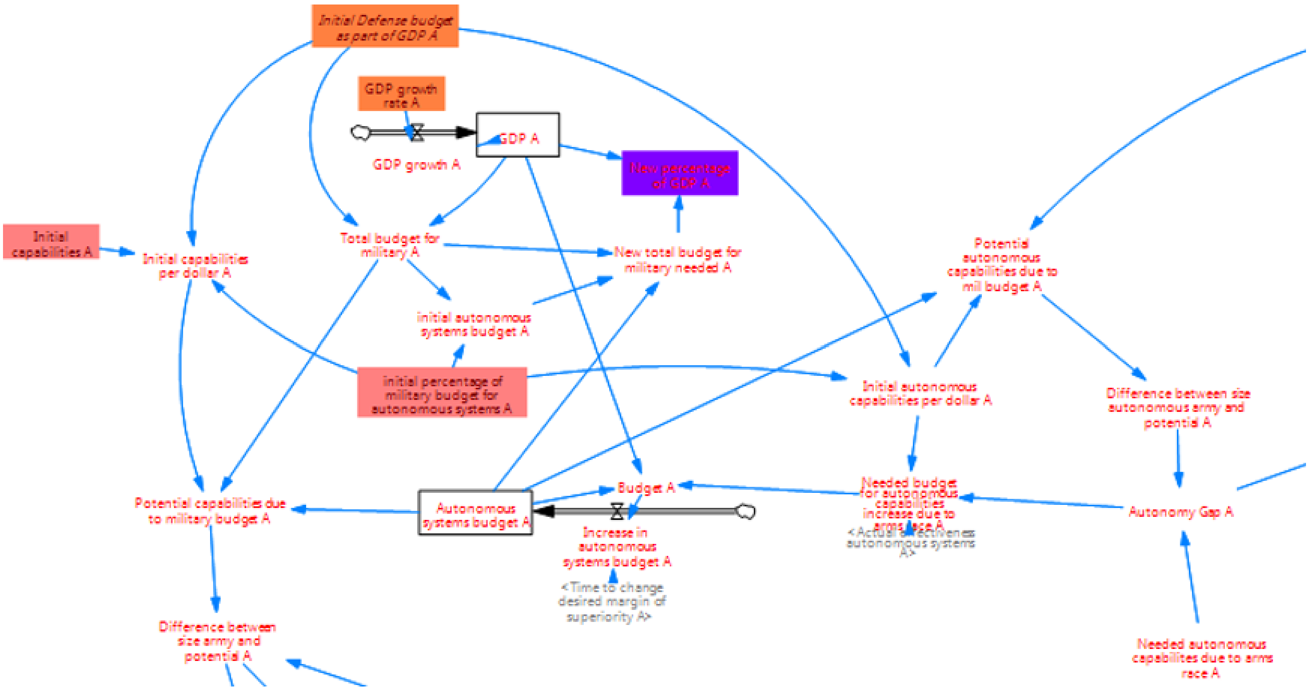

Initial Defense budget as part of GDP A
GDP growth rate A
GDP A
GDP growth A
New percentage of GDP A
Initial capabilities A
Initial capabilities per dollar A
Total budget for military A
New total budget for military needed A
initial autonomous systems budget A
initial percentage of military budget for autonomous systems A
Potential autonomous capabilities due to mil budget A
Initial autonomous capabilities per dollar A
Difference between size autonomous army and potential A
Potential capabilities due to military budget A
Autonomous systems budget A
Budget A
Increase in autonomous systems budget A
<Time to change desired margin of superiority A>
Needed budget for autonomous capabilities increase due to arms race A
<Actual effectiveness autonomous systems A>
Autonomy Gap A
Difference between size army and potential A
Needed autonomous capabilites due to arms race A

Figure 5: Budget for autonomous capabilities again only showing nation A.

If a nation’s budget suffices to meet its need in autonomous capabilities, there is no need to increase the available budget for autonomous capabilities. However, if the ”autonomy gap” grows, the percentage of GDP spent for autonomy and as such on defense needs to be increased. If the policy of restricting the funding for a nation’s autonomous military capabilities is implemented, then the nation cannot acquire new autonomous military systems.

Due to the arms race, there is a tradeoff to be made whether to spend available defense funds for more autonomous capabilities or for other military capabilities. A nation does not want to be behind with their autonomous capabilities and therefore with limited budget available, other military capabilities will be spent less on.

While autonomous systems are, as the word suggests, autonomous, there will always be to a varying extent a human crew around the autonomous systems (The Walking Soldier Podcast). These are operators and/or personnel for the maintenance of the autonomous systems. When planning where to allocate the available defense funds to, the costs that arise due to these personnel need to be considered (see Figure 6).

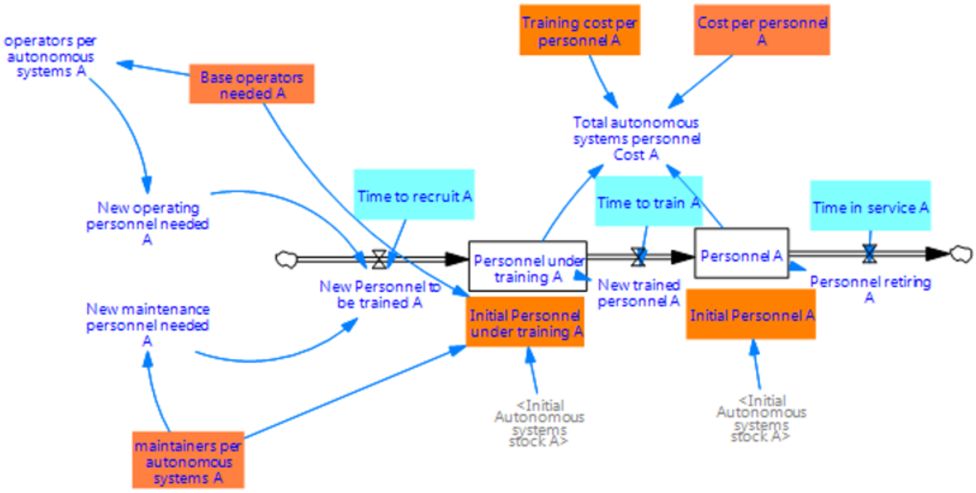

operators per autonomous systems A
Base operators needed A
Training cost per personnel A
Cost per personnel A
Total autonomous systems personnel Cost A
New operating personnel needed A
Time to recruit A
Time to train A
Time in service A
Personnel under training A
Personnel A
New Personnel to be trained A
New trained personnel A
Personnel retiring A
New maintenance personnel needed A
Initial Personnel under training A
Initial Personnel A
<Initial Autonomous systems stock A>
<Initial Autonomous systems stock A>
maintainers per autonomous systems A

Figure 6: Personnel for operating and maintaining the autonomous capabilities of Nation A.

Model uncertainties as well as parameter ranges can be found in subsubsection 3.4.2 in Table 1, Table 2 and Table 3. These values are derived based on literature, calculations and estimations made during discussions during interviews with experts.

3.4 Validation

After constructing a model, it needs to be determined whether it is fit for purpose, namely whether it can be used to answer the research question about the conditions under which arms races happen and how to influence them. For this, the model validation tests as described in section 2 are applied. This section thus reports on the results of these validation tests and concludes that the model is indeed fit for purpose within its defined scope. The code for this analysis can be found on my repository on GitHub.

3.4.1 Model Structure Tests

I tested the model structure in an iterative process, as explained in section 2. These tests include structure verification and parameter verification tests as well as dimensional consistency tests. To ensure boundary adequacy, I based the SD model as stock-flow structure on the dynamic hypothesis presented in the form of a qualitative causal loop diagram in section 2. This way I can ensure that the SD model captures the entire model scope and is fit for purpose, namely for identify conditions in which arms races happen and how to influence these. I did this process partly together with the consulted experts from subsection 2.2 which can be understood as the first iteration in Figure 2.

The three subsystems that are modelled in detail capture the elements a nation needs in order to build autonomous military capabilities: funding, skilled personnel and needed critical components (MoD, Concepts and Doctrine Centre, 2018). The race submodel captures the quantity of autonomous systems a nation want to acquire. Combined with the planning submodel, this captures the needed funding. Also, the race submodel captures the security dilemma, as here, the increase in one nation's security (in the form of more autonomous systems) decreases the perceived security of the other nations. In this submodel, arms races for quantity and thus for great numbers of capabilities arise. The technology subsystems models the development of the quality of the systems as a result from artificial intelligence and technological sophistication. This captures the need for skilled personnel and critical components. In this submodel, arms races for quality arise.

Economy and GDP is not modelled explicitly but stems from outside of the model boundaries. While this captures the scope of the model and ensures that it is fit for purpose, some aspects are missed: A domestic arms industry would profit from arms races leading to a growing economy. Also, trade with the autonomous capabilities (or civilian versions of these) can benefit the economy. Thus, arms races would lead to an increase in GDP. However, this does not affect the occurrence of arms races and the growing of tensions between two nations. Not including economy and the potential positive effects on the national GDP of the arms racing, makes the analysis more straight forward: arms races are represented by an increase of defense spending of 50% over 5 years. Defense spending is measured as percentage of GDP. If this GDP grows due to the arms racing, some arms racing scenarios might not count as arms races as they increase the GDP and by this decrease the defense spending as percentage of this GDP.

Scientific research is not modelled explicitly and happens outside of the model boundaries. This assumes that most of the research into autonomy and robotics is not driven by military demand, but also by civilian demand. Thus, access to skilled personnel and to critical components in the mode cannot be influenced by the nations. A limitation of this is that the model does not capture measures that nations can take to influence their own access to these two factors. However, the purpose of the model is not to analyze how a nation can efficiently build autonomous military capabilities. On the contrary, the purpose is to analyze the occurrence of arms races. For this, taking scientific research as exogenous does not limit the model's application.

Not included in the model are the cultural differences between nation A and nation B. These might have an influence on the perception of threat, of trust and on communication between the two nations. These mechanisms are represented in the model as bias factors. These bias factors are uncertainties in the model for which I experiment with different value ranges. For instance, a very high bias to perceive the opposing nation's stock in autonomous capabilities represents a nation that has very little trust into its opponent. Thus, by experimenting with different values for these biases, I can assure that the model is fit for purpose for analyzing arms races without including culture.

In the model, I do not include diplomacy. This, as well as economy and trade, represent power instruments a nation has to pursue its objectives. In the model, nations have different strategies to influence arms races. These include bilateral agreements as well as targeting the opposing nation. For applying these strategies, nations need means in order to do so and to be successful. In the current implementation of the model, the nations do not have these means. Instead, whatever strategy the nations pursue, is successful. Still, this fits with the scope of the model, which is not about modelling a state's power and how to apply it, but to analyze the conditions for arms races. I do investigate how nations can influence arms races. For this, the model is valid under the assumption that nations are successful in applying their respective strategy.

I do not model other disruptive technology. While a nation will have to trade-off between how much it invest into which technology, other disruptive and emerging technology can have an influence on the available funding for autonomous capabilities. However, other military capabilities are included in the model such that I capture this trade-off to some extent. At the same time, any other disruptive military technology that a state invests in could result in its own arms race. Thus, this does not necessarily influence the dynamics of arms races for autonomous military capabilities. It might influence the exact amount of funding that goes into autonomous capabilities, but not their behavior over time. Thus, the model is valid, as long as there is not a technology that proves to be more of a game changer than autonomous capabilities such that a nation has no interest in developing these anymore.

Other military engagement or a war is not included in the model. If there was a war between the two nations, the defense spending and thus the available funding for autonomous capabilities would be completely different. Thus, the model is only valid if there is no actual military conflict by any of the two modelled nations.

In the model, I do not implement internal stability within the two nations. Similarly, to above, if there was a high degree of instability (e.g., a civil war going on), a nation will have different priorities than engaging into an arms race with another nation. Hence, the model is only valid if a nation is more or less stable.

I do not model third parties that could figure as strategic partners or as mediators between two nations in tensions. One nation with a very strong partner would be more powerful than a nation without. Since I conduct many experiments with large parameter ranges over all model uncertainties, the model structure is still valid to make conclusions about a potential third party. Thus, I can use the model to conclude about which strategies third parties can employ to influence arms races, depending on their objectives to either prevent arms races or to have one nation win over the other.

3.4.2 Model Behavior Tests

With the model structure validated, with this section I report on the model behavior tests that are introduced in section 2. The results of these tests suggest that the model is fit for purpose and can, within its scope, be used to answer the research questions about the conditions for arms races and how to influence them.

Behavior sensitivity Tests in the form of open exploration of the model shows what behavior the SD model can generate. Therefore, 5000 experiments are performed with the EMA workbench (Kwakkel, 2022) using Latin Hypercube sampling over the uncertainty space (see subsubsection 3.4.2 in Table 1, Table 2 and Table 3 for their plausible ranges) without any active policies. In Vensim, Euler is used as integration method with a time step of 0.00390625 Years of a duration of 30 years. The results on the extreme condition tests can be found in Appendix C.

Two outcomes of interest are studied: On the one hand, the defense spendings of hegemon nation A and rising nation B is analyzed over the course of the simulation time, as a percentage of their GDP. On the other hand, the size of the nations' autonomous capabilities and its dynamic over time are analyzed. To begin with, the defense spending of nation A and B can be seen in Figure 7. This scatterplot shows the final time step, namely year 30 at the end of the simulation time. The correlation of nation A's final defense spending with nation B's final defense spending can be seen here. A somewhat linear relationship can be observed between the expenditures of both nations, especially at the higher values. Due to the model structure, the defense spending grows as a response to a nation's need to increase its autonomous capabilities and to increase the size of its personnel.

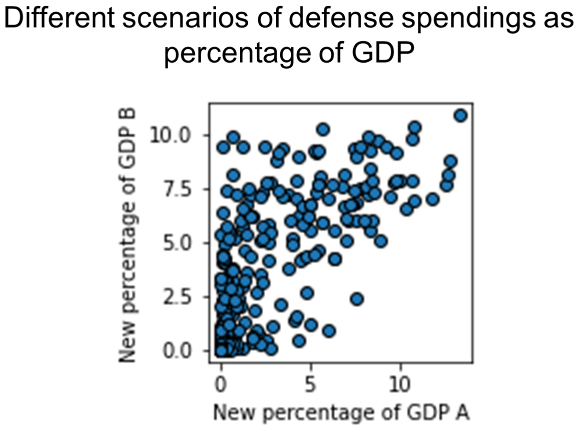


Figure 7: Defense spending as percentage of GDP at the final time step in year 30 of hegemon nation A versus rising nation B. Each point in this scatterplot represents one scenario.

Outliers can also be seen in Figure 7 and are defined as scenarios with an extreme defense spending, exceeding 46% of a nation's GDP (Collier and Hoeffler, 2002). Of all experiments run, about 29% can be determined as extreme. While in this scatterplot only the final defense spending values can be seen, expenditures of more than 46% of GDP also happen at other time steps, as can be seen later in Figure 9. In these scenarios, hegemon nation A is initially technologically advanced to rising nation B. However, during the simulation time, B overtakes A in the autonomy of the military capabilities as can be seen in Figure 8. Moreover, in these scenarios rising nation B needs less personnel to maintain its autonomous capabilities, nation A soon is recruiting way more personnel than nation B. At the same time, A's GDP is greater, while it is growing slower, due to which A has greater overall military capabilities. Hence, while the autonomous capabilities are less technologically sophisticated after B has overtaken A in its autonomy, B still drives the arms race to keep up with the strength of A, given by its personnel and other military capabilities, leading to exploding defense spendings as can be seen in.

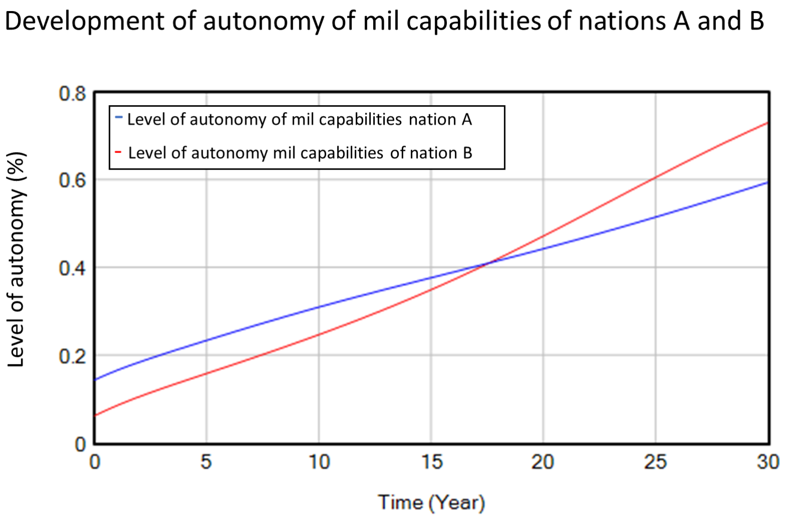


Figure 8: Level of autonomy of military capabilities when defense spendings explode. In this specific scenario, rising nation B overtakes hegemon nation A around year 17.

The dynamic behavior over time of the defense spendings of nations A and B is also interesting to consider. A difference in temporal behavior can be observed between those scenarios that result in extreme defense expenditure and those that do not. The extreme defense spendings behave in two different ways (see Figure 9). In the first case, expenditure grows in a straight line that starts at a specific time step and increases until the end of the simulation time. In the other case, defense spending starts to increase at a certain time step and plateaus, resulting in an S-shaped growth of military expenditure. Comparing these observations to Figure 10, it can be seen that the realistic defense expenditure behaves in four different ways. First, there are the cases with no change in defense spending, and second, where the expenditure decreases slightly. Third, there are those scenarios where the defense spending oscillates and finally those, where it increases steeply towards the end of the simulation time. All the extreme cases are arms races, where the defense spending of at least one nation increases by more than 50% over 5 years. In total, about 40% of all scenarios result in arms races.

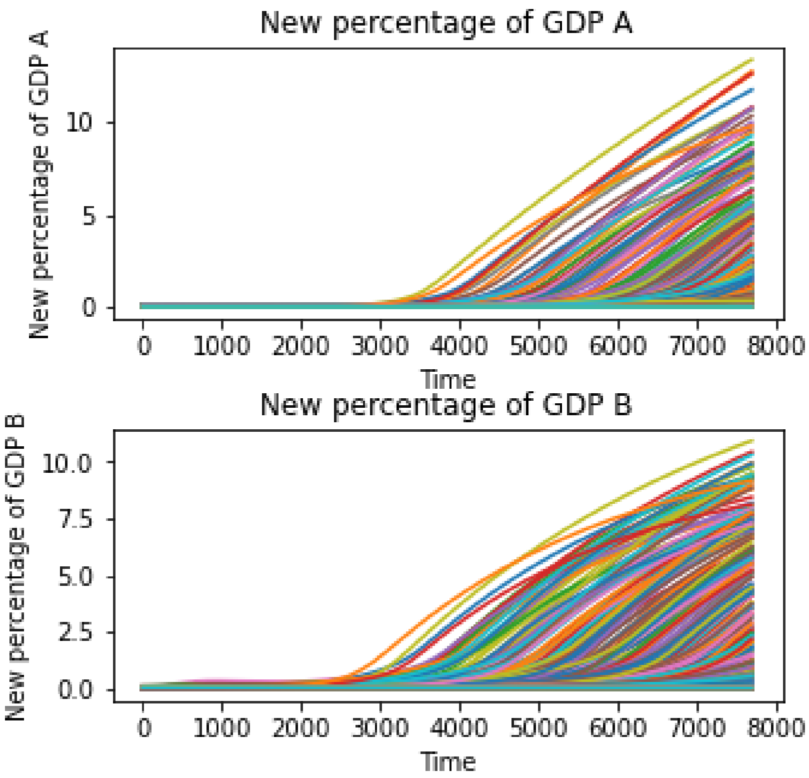


Figure 9: Temporal behavior of defense spending as percentage of GDP for nation A (top) and nation B (bottom) over the simulation time of 30 years.

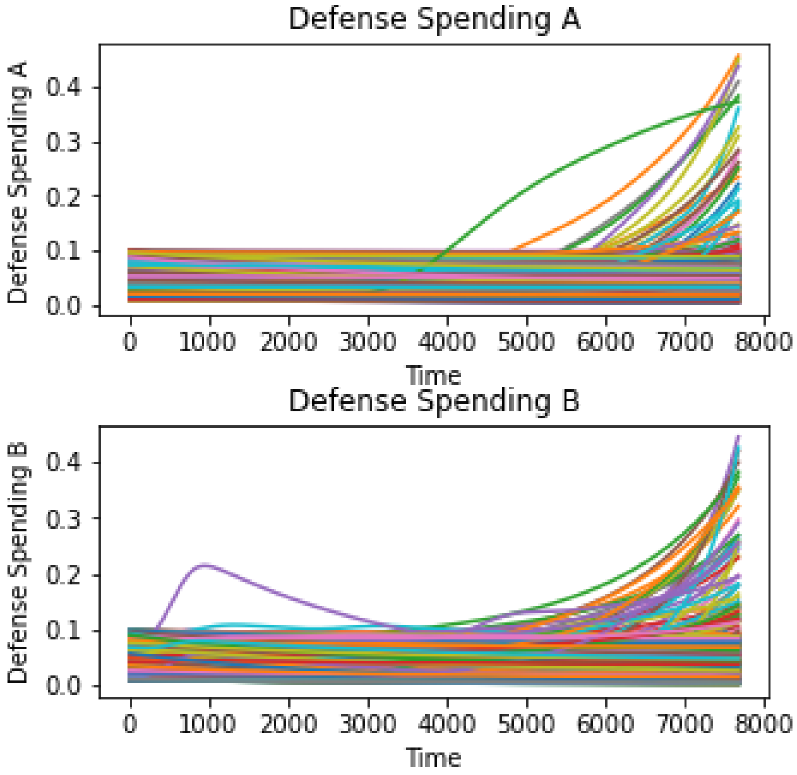


Figure 10: Temporal behavior of defense spending as percentage of GDP for nation A (top) and nation B (bottom) over the simulation time of 30 years without the outliers.

Furthermore, the stocks of piled up autonomous systems of nation A and B are another outcome of interest to investigate the occurrence of arms races. A nation varies this stock in response to the opposing nation's autonomous capabilities after a certain reaction time (given by the parameter "Time to change desired margin of superiority") has elapsed. A closer look at this one case reveals the specifics of why this leads to such an extremely high amount of autonomous military capabilities and hence related exploding defense spendings. The exploding number of autonomous capabilities is driven by different mechanisms on both sides, as a result to the parameter setting of this scenario. As can be seen in Figure 11, the capabilities of hegemon nation A achieve full autonomy, while the ones of rising nation B never do. At the same time, nation A needs very little personnel to operate and maintain its autonomous capabilities while nation B needs quite a large number of maintainers. While in principle it would not need a lot of operators, it does after all need to recruit a lot of personnel of its autonomous capabilities as their level of autonomy are rather low. Also, nation B needs a great number of personnel to maintain its autonomous capabilities. In addition, due to nation A's GDP being very high, it has more means in general to finance its defense expenditure. Hence, nation A can initially still increase its investments into other military capabilities regardless of its rather increased percentage of the military budget going to autonomous capabilities. Nation B however has to trade-off its other military capabilities for autonomous capabilities from the very beginning in order not to lose the arms race.

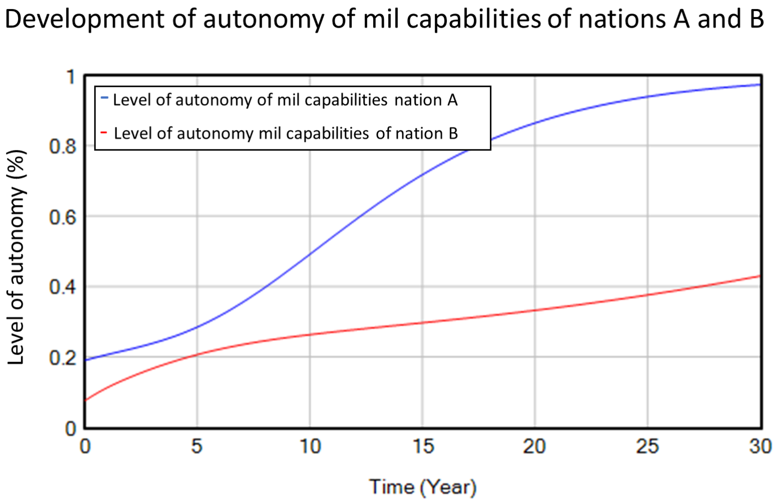


Figure 11: Level of autonomy of military capabilities when the stock of autonomous military capabilities explodes. In this specific scenario, hegemon nation A reached full autonomy around year 17, while rising nation B always lacks very much behind.

This outlier is happening, because nation A, with its huge GDP, has a great capacity for investing into both autonomous and other military capabilities. At the same time, the desire for nation B to heavily invest into autonomous capabilities is driven by nation A's technological superiority and its greater overall military capabilities. In fact, rising nation B even overtakes nation A in term of quantity at some point. Hegemon nation A then reacts to nation B's increase in autonomous capabilities by increasing its own stock. This is not only due to the growing stock of B's autonomous capabilities, but also due to B's related increase of recruiting personnel for operating these systems.

The validity of the model is still given, regardless of the lone outlier where the autonomous capabilities for both nations explode, and of those scenarios where the defense spending explodes. On the one hand, to conclude on this scenario, it can be debated whether this constitutes a realistic parameter combination. It is questionable whether a great power competition between nations like A and B in this scenario would come about. Most probably, a nation such as nation B in this scenario would not attempt to compete in a race towards autonomous capabilities against a nation A when lacking this much behind in terms of their development of autonomy. In fact, starting from year 8 in the simulation, nation A is achieving a level of autonomy that is more than twice as high than nation B. At this point in time, a nation would probably reconsider contesting the status quo established by nation A. At the same time, the parameter combination of needed personnel might not be tenable in real world. While the level of autonomy of B's capabilities is relatively low, it chooses to recruit very little personnel for the operation of its fleet while recruiting a large amount of personnel for the maintenance of its fleet. When assuming that nation B would indeed opt for more operators for its autonomous capabilities, the gap between the sizes in personnel between A and B might not be as large and hence B would not need to drastically increase its personnel. This would prevent the additional tension arising from racing against each other's size in personnel.

On the other hand, even though the resulting defense spendings cannot be taken as real values, the scenarios under which these expenditures take place can still give an indication about when a competition between two nations such as A and B could escalate. The exactness of the numbers that the model generates depends on the exactness of the uncertain parameter values and parameter ranges. As my resources for finding accurate values for these parameters are limited, these are based on assumptions, estimations, calculations, and discussions with experts. Consequently, the model does under certain circumstances generate extreme defense spendings and those values need to be taken with a grain of salt. However, the model is still deemed valid and fit for purpose. The purpose of the model is namely not to accurately predict how much an arms races would cost, but to identify the conditions under which arms races between hegemon nation A and rising nation B occur and what can be done about it.

Model Uncertainty	Min Value	Max Value	Unit
Initial autonomous systems stock A	1000	5000	Capability
Initial autonomous systems stock B	100	1500	Capability
Sensitivity to opposing autonomous systems production A	0.75	1.25	Dmnl
Sensitivity to opposing autonomous systems production B	0.75	1.25	Dmnl
Average life of autonomous systems A	20	30	Year
Average life of autonomous systems B	5	25	Year
Time to change desired margin of superiority A	0.7	2	Year
Time to authorize autonomous systems development A	0.7	2	Year
Time to change desired margin of superiority B	0.7	2	Year
Time to authorize autonomous systems development B	0.7	2	Year
Time to perceive opposing autonomous systems A	0.25	0.75	Year
Time to perceive opposing autonomous systems B	0.25	0.75	Year
Sensitivity to opposing autonomous systems production A	0.75	1.25	Dmnl
Bias in estimating opposing autonomous systems A	0.75	1.25	Dmnl
Sensitivity to opposing autonomous systems production B	0.75	1.25	Dmnl
Bias in estimating opposing autonomous systems B	0.75	1.25	Dmnl
Bias in estimating domestic autonomous systems A	0.75	1.25	Dmnl
Bias in estimating domestic autonomous systems B	0.75	1.25	Dmnl

Table 1: Summary of plausible parameter ranges of the racing submodel

Model Uncertainty	Min Value	Max Value	Unit
Initial tech sophistication under development A	0.25	0.3	Technology
Initial tech sophistication under development B	0.2	0.25	Technology
Initial level of autonomy under development A	0.25	0.3	Technology
Initial level of autonomy A	0.1	0.2	Technology
Initial level of tech sophistication B	0.01	0.1	Technology
Initial level of autonomy under development B	0.2	0.25	Technology
Initial level of autonomy B	0.01	0.1	Technology
Initial level of tech sophistication A	0.1	0.2	Technology
weight autonomy B	0.5	1	Dmnl
weight autonomy A	0.5	1	Dmnl
v tech A	1	20	Dmnl
v tech B	1	20	Dmnl
Q tech A	1	20	Dmnl
Q tech B	1	20	Dmnl
v AI A	1	20	Dmnl
v AI B	1	20	Dmnl
Q AI A	1	20	Dmnl
Q AI B	1	20	Dmnl

Table 2: Summary of plausible parameter ranges for the technology submodel

Model Uncertainty	Min Value	Max Value	Unit
Initial capabilities A	100000	200000	Capability
Initial capabilities B	10000	150000	Capability
Average lifetime of capabilities A	20	30	Year
Average lifetime of capabilities B	5	25	Year
Capability procurement speed B	7	10	Year
Average procurement delay other cap B	7	10	Year
Capability procurement speed A	2	7	Year
Average procurement delay other cap A	2	7	Year
Average procurement delay A	2	10	Year
Average procurement delay B	2	10	Year
GDP growth rate A	0.01	0.07	1/Year
Initial GDP A	2E+12	2E+13	Dollar
Initial Defense budget as part of GDP A	0.01	0.05	1/Year
Initial Defense budget as part of GDP B	0.01	0.05	1/Year
GDP growth rate B	0.05	0.1	1/Year
Initial GDP B	8E+11	2E+12	Dollar
percentage of military budget for autonomous systems A	0.001	0.9	1/Year
percentage of military budget for autonomous systems B	0.001	0.9	1/Year
Base operators needed A	0.5	10	Personnel/Capability
Base operators needed B	0.01	2	Personnel/Capability
Maintainers per autonomous system A	0.01	5	Personnel/Capability
Maintainers per autonomous system B	0.01	5	Personnel/Capability
Time to recruit A	0.5	2	Year
Time to recruit B	0.5	2	Year
Time to train A	1	5	Year
Time to train B	1	2	Year
Time in service A	7	25	Year
Time in service B	5	25	Year
Training cost per personnel A	5000	25000	Dollar/Personnel
Training cost per personnel B	1000	10000	Dollar/Personnel
Cost per personnel A	50000	250000	Dollar/Personnel
Cost per personnel B	10000	100000	Dollar/Personnel

Table 3: Summary of plausible parameter ranges of the planning submodel

In subsubsection 3.4.2 in Table 1, Table 2 and Table 3, I summarize the parameter ranges for the model uncertainties. I derived many of these based on experimentation and calculations. I derived the defense spendings, training costs, GDPs and initial stocks of military capabilities based on data from the defense spendings of nations with varying GDP sizes. For parameterizing the technological developments (parameters V tech, v AI Q tech and Q AI), I adapted a sigmoid function until the expected behaviors were modelled. Time delays, procurement delays, training times, lifetimes of capabilities are estimations. The same goes for biases and sensitivities: I assume that a nation either underestimates or overestimates its opponent by around 25%. Personnel needed per autonomous system is also an estimation.

As part of the model validation process, I analyzed the model structure and the model behavior to determine whether the model is fit for purpose. In this section I describe the results of these tests performed in the form of open exploration using the EMA workbench. The model can produce a wide range of different behaviors of defense spendings. While a subset of these scenarios do not result in realistic outcomes, I can still use the model for the subsequent analysis of under which circumstances arms races might happen and how nations could influence these.

3.5 Experimental Setup

I connected the SD model that I constructed in Vensim with the EMA workbench which I use for applying PRIM and RDM to the model. For scenario discovery, I perform 5000 experiments using Latin Hypercube sampling over the uncertainty space (see Table 1, Table 2 and Table 3 in subsubsection 3.4.2 for their plausible ranges) without any active policies. With these parameter settings, I characterize hegemon nation A as the stronger nation with a head start both in quantity and quality in autonomous military capabilities with the higher model ground. Rising nation B is the opposite. The code for this analysis can be found on my repository on GitHub.

I use Euler as integration method with a time step of 0.00390625 years of 30 years. For identifying arms races, I conduct 5000 experiments and for comparing policies I conduct 1000 experiments per strategy over the uncertainty space. I use PRIM with a lenient objective function at the default settings of the EMA workbench with Latin Hypercube sampling. I define outcomes of interest are defined as those scenarios where arms races are happening. To measure the performance of the different strategies, I use regret as robustness metrics. Hence, I compare all the different strategies and determine the one performing the best, where the best is the one with the least number of arms races over the uncertainty space. After identifying arms races, I cluster these based on their dynamic behavior over time. For clustering the outcomes of interest, I used the complex invariant distance between the experiments, producing 3 clusters.

4 Results

This section reports on the results of conducting scenario discovery and robust decision making as explained in section 2 to answer the research questions: under which conditions do arms races occur and how to influence them. First, the experimental setup of the Vensim model and the EMA workbench are introduced. Then, subsection 4.1 introduces the results of the discovery of arms races. Characteristics of arms races are identified using the PRIM algorithm and by determining parameter values and combinations that entail arms races. Last, subsection 4.2 reports on the results from applying different policies to influence the number of arms races. Scenarios in which these policies do and do not well are also described.

4.1 Identification of Arms Races

Arms races are defined as scenarios where the defense spending of a nation increases by more than 50% over a period of 5 years (Diehl, 1985). This reflects the second condition of Caspary (1967) for characterizing an arms race; namely that the arming is a burden on economy. There are around 40% of the experiments that result in arms races. Two third of these arms races happen at extreme defense spendings where at least one of the nations uses more than 46% of their GDP for defense expenditure (Collier and Hoeffler, 2002). At the same time, all extreme defense spendings are arms races. Most races, this being about 79%, are bilateral races, meaning that they are driven by both countries increasing their defense spendings at the same time. There are also some unilateral races, where either only hegemon nation A increases its defense spending by more than 50% over a period of 5 years (about 9%), or only rising nation B does (about 13%). Finally, the winner of arms racing scenarios more often than not is rising nation B, while in no arms racing scenarios, hegemon nation A results as the winner from the competition for autonomous military systems.

4.1.1 Discovering Arms Races

Arms races can be characterized by 4 parameters as can be found using the PRIM algorithm. The combination of these parameter ranges describe arms races. To summarize the results from PRIM, arms races are thus caused by shorter response times to each other's arming actions (i.e., small value for the parameter ”Time to change desired margin of superiority” for A and B) and by nation B has a relatively small size of initial (overall military) capabilities. In addition, the bias in estimating opposing autonomous capabilities B is slightly higher. Due to the limitations of PRIM (see Appendix D), these are representative of only 25% of the found arms races. Hence, the next section dives more into which uncertainties are responsible for producing arms races.

4.1.2 Characterizing Arms Races

To get a deeper understanding of which parameter combinations lead to arms races, additional analyses after PRIM are conducted. For this, scenarios with arms races and without arms races are separated. For each parameter in the uncertainty space, the average is then taken once for arms racing scenarios and once for scenarios without arms races. Then, the relative difference between both cases is taken. The result is summarized in Figure 12. This additional step after performing PRIM is necessary since the PRIM algorithm was only able to characterize about 25% of the scenarios (see Appendix D).

First, arms races happen where both hegemon nation A and rising nation B take a shorter than average amount of time to change their desired margin of superiority. This means that there are quicker to responding to each other than in not arms racing scenarios: As soon as for instance nation A perceives that nation B is increasing its stock of autonomous military capabilities, nation A decides, after a very short amount of time, to invest more into autonomous capabilities (and vice versa). This finding is also supported by the results of PRIM as was introduced above.

Second, in arms races, hegemon nation A on the one side is characterized by having both a larger than average growth rate of its GDP and initially already spends more funds on autonomous military capabilities. Due to the combination of these factors, A has, in the arms racing scenarios, an even greater (military) power than in the average scenarios. Rising nation B on the other side is characterized by a smaller initial stock of military capabilities in general as well as of autonomous capabilities. At the same time, nation B recruits less in personnel for operating and maintaining its autonomous capabilities. Hence the military power of B is even lower in arms racing scenarios than on average, due to its smaller capability stocks and its smaller personnel stock. These findings are also to some extent covered by using the PRIM algorithm.

Last, the development of AI and of autonomy of rising nation B is faster than in not arms racing scenarios. Hence, the autonomy of the autonomous capabilities of B increases faster. Due to this, even though B has a smaller initial number of autonomous capabilities, its autonomous capabilities increase faster in autonomy, earlier than those of hegemon nation A.

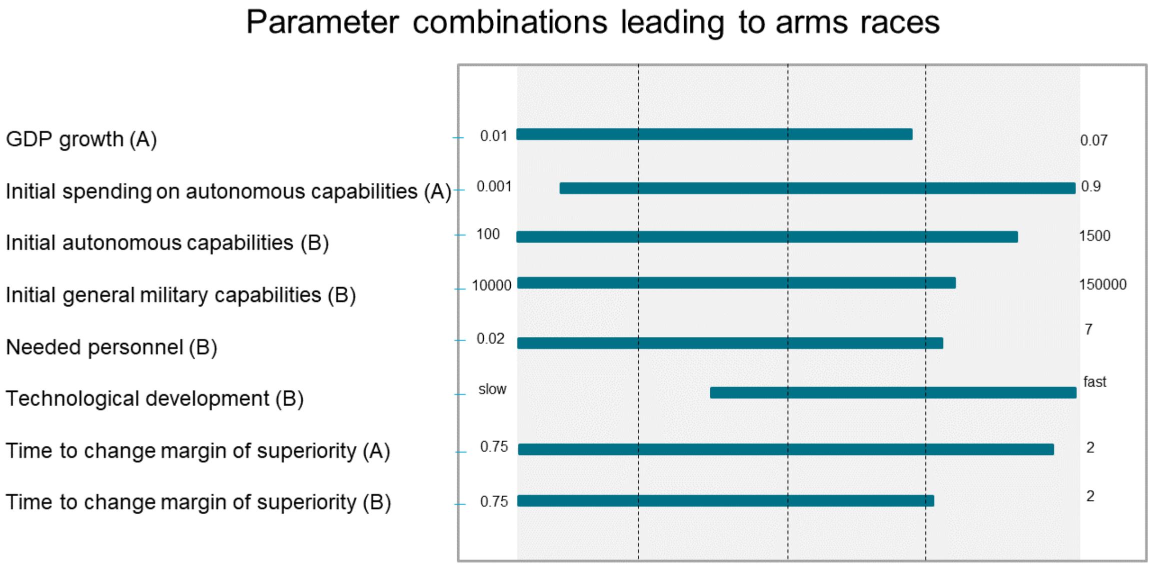


Figure 12: Characterization of arms races.

4.1.3 Different Dynamics of Arms Races

Discovering arms races shows that defense spending and hence arms races behave in different ways: They either oscillate, reach a plateauing in an S-shaped growth, or increase steep increase. There are also the extreme arms races that were already discussed in subsubsection 3.4.2. Plus, there are the unilateral and the bilateral arms races. In addition, there are two types of unilateral arms races; the ones that are driven by hegemon nation A and the ones that are driven by rising nation B. The conditions for the latter are summarized in Figure 13. To get an understanding of the different temporal behaviors of arms races that are observed in different scenarios, these are once clustered based on the dynamic behavior of the defense spending of nation A and once based on the one of nation B The results of this analysis can be found in Appendix E.

The unilateral races can be compared to bilateral races to determine the reason why the respective opposing nation does not get involved in the arms racing. On the one side, in the unilateral arms races that are only driven by rising nation B (14% of all arms races), nation A has an extremely high initial investment into autonomous capabilities, while this investment for B is lower than for bilateral arms races. Interestingly, while the initial general capabilities of nation B are lower than in bilateral arms races, its initial autonomous capabilities are higher.

On the other side, in the unilateral arms races that are only driven by hegemon nation A (9% of all arms races), the conditions as described above are to some extent reversed: nation B has an extremely high initial investment into autonomous capabilities, while this investment for A is lower than for bilateral arms races. In addition, the initial autonomous capabilities of nation A are higher and nation B's initial general and autonomous capabilities are lower. Interestingly, the GDP growth rate of nation A is in fact higher than in bilateral arms races.

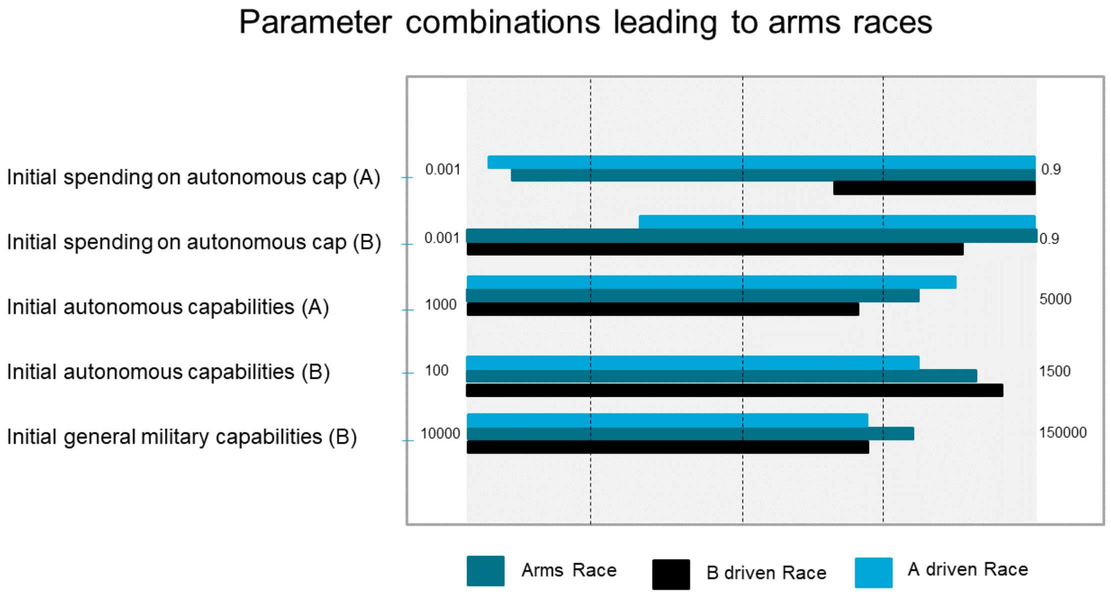


Figure 13: Characterization of different types of arms races.

4.1.4 The Outcome of Arms Races

After analyzing different behaviors of arms races, I also evaluate which nation wins the arms race under which conditions. This is important as the evolution of the defense spending over time as percentage of GDP is only one chosen indicator of measuring arms races. This indicator on its own cannot give an indication on which nation would eventually end up with the superior autonomous capabilities. Therefore, as I explain in subsection 2.3 a nation's weighted quantity is computed: For this, at the end of the simulation time, namely year 30, the number of systems a nation has accumulated it multiplied (and hence weighted) with their overall effectiveness.

Type of Race	A wins
Bilateral	22 %
A driven	24 %
B driven	32 %
No Race	50 %

Table 4: Percentage of arms races that hegemon nation A wins, depending on the different types of arms races, when no policy is implemented.

Hegemon nation A can win the competition for autonomous military systems with nation B, when there are no arms races going on (see Table 4). In fact, as soon as there is an arms race, chances of nation A winning are rather low. The scenarios in which nation A does win the arms race after all are characterized by nation A having a faster procurement for autonomous systems while B having a slower procurement, next to all other characteristics of arms races as already described above. On the contrary, in scenarios where nation B wins, nation A generally needs even more people for the operations of its autonomous capabilities again reflecting the differences in ethical standards between the two nations. These parameter combinations can be seen in Figure 14.

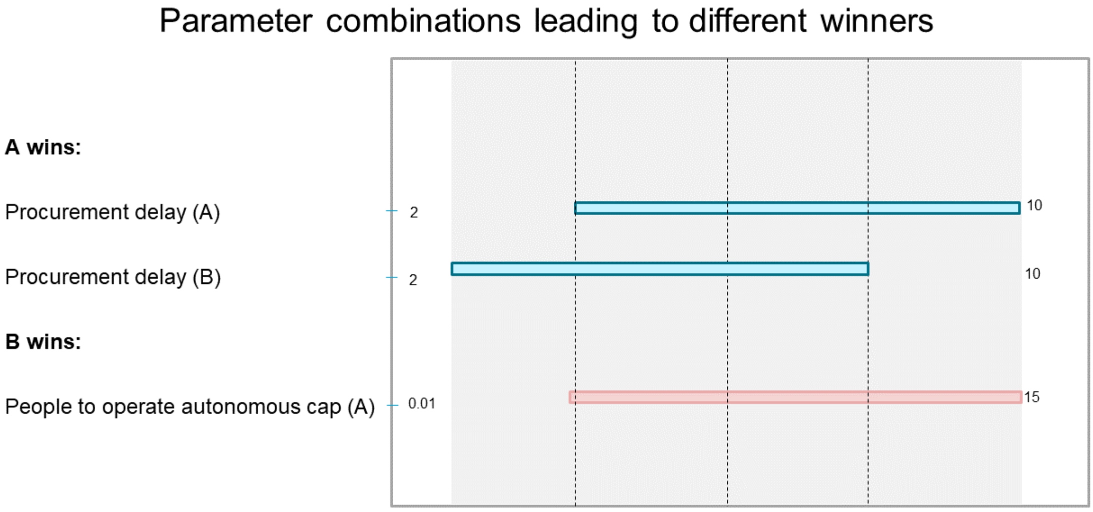


Figure 14: Conditions for winners, on top of regular arms racing conditions, where the parameter ranges in blue denote arms races that nation A wins and in red those that nation B wins.

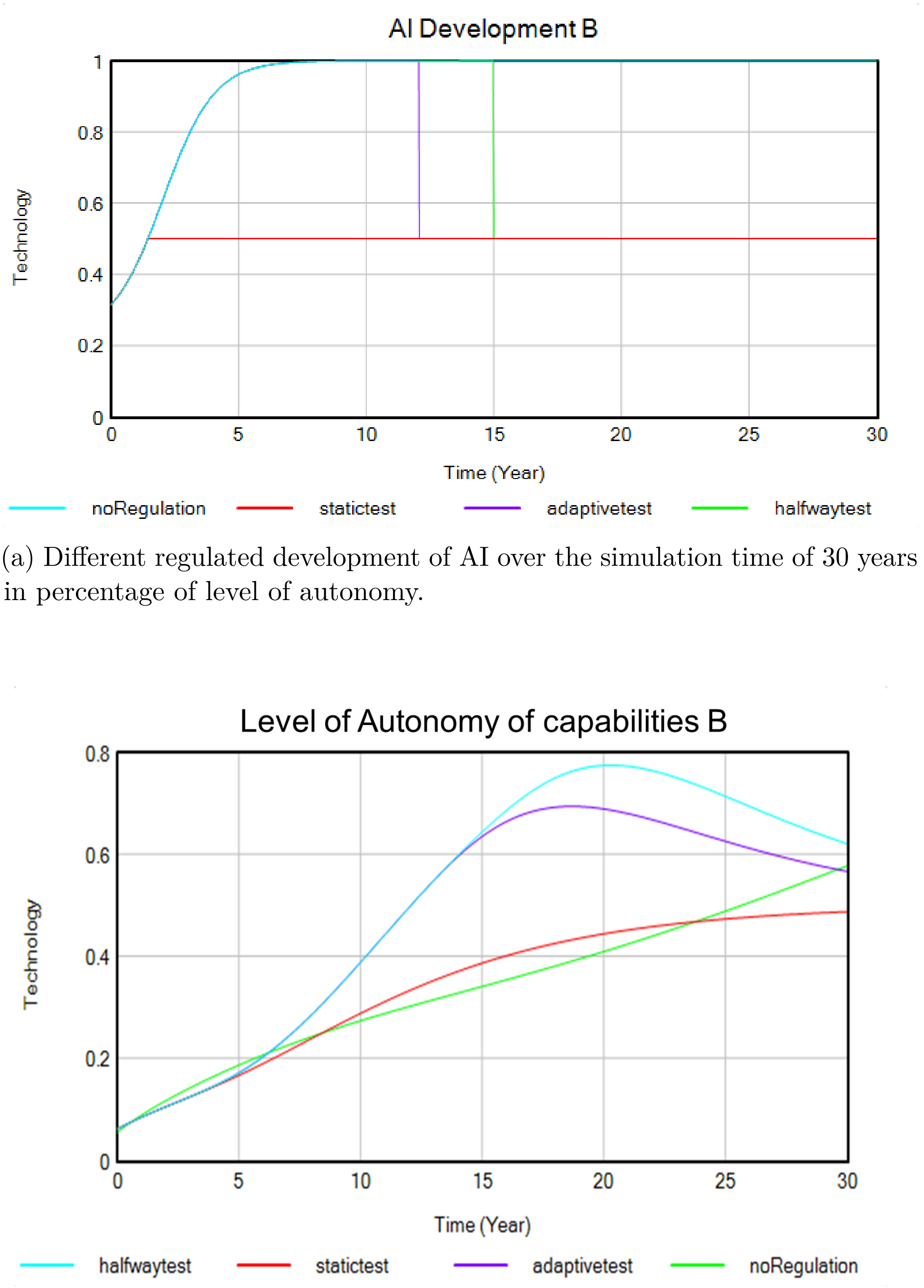
4.2 Strategies to influence Arms Races

After characterizing arms races, strategies to influence these are needed. For this, a distinction can be made between a quantitative and a qualitative arms race (Intriligator and Brito, 1984), which enables different possible points for a nation to address the occurrence of arms races. Further, there are two different options of strategies for a nation. On the one hand, to prevent arms races from happening, nations can restrain themselves from engaging into a competition. This relates to regulation strategies for addressing the qualitative arms races, where hegemon nation A and rising nation B bilaterally agree on a level of autonomy of their autonomous military capabilities, that no nation can surpass. On the other hand, a nation can aim to retain the upper hand themselves: for this, it target the opposing nation and impedes it in its efforts in the race for autonomy. Nation A can thus aim at influencing B's efforts in investing and developing autonomous military capabilities, which addresses both quantitative and qualitative arms races.

4.2.1 Regulation Strategies

A cap on autonomy will be implemented with the regulation strategies. This cap of a level of autonomy of 5 when using the scale of Sheridan and Verplank (1978) refers to the system being able to make and execute its own suggestions after human approval. This can be seen as "meaningful human control" (Ekelhof, 2018). To begin with, the difference between a nation's development of AI and the actual level of autonomy of the nation's autonomous military capabilities needs to be clear. With further development of AI over time, the level of autonomy of the autonomous military capabilities will eventually increase, after some time delays. The development of AI is measured by the potential level of autonomy that can be reached. However, due to time delays, they are not the same. The development of AI might have reached a potential for level 5 autonomy. However, due to the time it takes to implement this into the acquired systems, there is a delay until the actual level of autonomy of the autonomous capabilities is also of 5. The regulations aim at restricting the actual level of autonomy of a nation's systems by restricting a nation from developing AI, as I explain in section 3.

There are three ways in which the regulation on the level of autonomy is implemented. In the first case of the adapting strategy, as soon as one nation attains a level of autonomy that is equal or higher to 5, both nations will not be allowed to further develop their AI to achieve levels of autonomy higher than 5. The second and midway strategy restricts the development of AI to a level of 5. This means that before this point in time the nations can have levels of autonomy higher than 5, but starting year 15, they need to decrease their levels of autonomy. The last and static strategy places a cap on autonomy from the very beginning of the simulation time. Here, starting year 0, no nation is ever allowed to increase the potential of their AI to achieve levels of autonomy of their capabilities to higher than 5. The different regulations are shown in Figure 15.



(a) Different regulated development of AI over the simulation time of 30 years in percentage of level of autonomy.

(b) Percentage of level of autonomy over simulation time of 30 years as a result of regulated development of AI.

Figure 15: Effect of different regulations on the development of AI and the resulting level of autonomy for nation A over the simulation time of 30 years.

The best results, namely the least number of arms races where hegemon nation A wins most often is obtained when applying no regulation strategy. When running the model with a regulation implemented, hegemon nation only wins the competition a third of the times and thus loses twice as often as without any policy. At the same time, however, there will always be around 30% less arms races than without any policy. The best performing strategy in terms of reducing the number of arms races is when applying regulation starting year 0 of the simulation, as can be seen in Figure 20. Here, the number of arms races can actually be reduced by 40%.

The best performing strategy for having hegemon nation A win the competition is when applying the regulation only halfway through the simulation at year 15. Then, nation A is able to win in 33% of the scenarios.

Scenarios in which arms races with policies take place are analyzed using PRIM (see Appendix D for the implementation of the latter). First, the static policy which implements a cap on autonomy beginning year 0, can be described with lower initial (general military) capabilities of nation B as well as a shorter time to change desired margin of superiority for nation A and a slightly smaller initial percentage of military budget for autonomous systems B. To explain the slightly more arms races happening with the static policy than without any policy, it needs to be noted that here, the reaction time of nation A is shorter than for the no policy arms races and the initial capabilities of nation B are larger than with no policy. Hence, since A is more reactive in these scenarios than compared to the no policy ones, it engages in arms races sooner and more often. At the same time, as B has a slightly greater military power due to higher initial capabilities as on average, A sees it more often as threatening the status quo. While this is as such not very different to the scenarios without policy implemented, the main difference is that there is no more the possibility of outrunning the opposing nation with the quality of the autonomous capabilities. Before, with no policy implemented, there was the option of reducing the perceived threat by opting for the quality versus the quantity option. Now, the quality option does not exist anymore since at some point, when both nations have reached level 5 autonomy, they can only compete in terms of numbers for which they need to increase their budget.

Second, the adapting policies can also be described with the same uncertainties. Here, the response time of nation A is longer, even longer than when no policy is implemented. Still, the initial capabilities of nation B are a bit higher than without policy. Hence, there are more arms races than without policy since B is seen more as a threat in A's perception. Thus, there are more arms races than without any policy, but less than with the static policy, as A is not as reactive. Plus, the option for competing in terms of quality is not taken away completely. The two nations are only restricted from developing their AI from that moment on where one nation has reached level 5 autonomy. As can be seen in the specific scenario in Figure 15a, when implementing the adapting policy in year 24, the potential of AI had already reached a possible level of autonomy of the autonomous capabilities of 7. Hence, due to the time delays in the system, and before the average level of autonomy has decreased again to a level 5, the two nations still have the option in competing in terms of quality. For this, they do not need to excessively increase their defense spendings.

Last, the midway policy is also characterized by low initial capabilities for nation B and a shorter response time for nation A (longer than the no policy and static policy arms races though), plus a higher GDP growth rate A. At the same time, the initial capabilities of B are even higher than with any other regulation and the response time for nation A is the same as for the static and the adapting policy, thus a bit higher than without any policy. Overall, this policy leads to more or less the same results as the static policy. This can be explained by looking at Figure 15a and at Figure 15b.

4.2.2 Nation A influences nation B

Three different strategies can be employed by a nation to influence arms races. To begin with, the outcomes of only having hegemon nation A target rising nation B are analyzed. Nation B has no means of influencing nation A.

First, with the funding strategy, nation A targets the quantity of autonomous capabilities that nation B can achieve. It does so by preventing nation B from access to funding for building autonomous capabilities. This happens as soon as nation B will have the same quantity of autonomous systems than nation A. Then, without the needed funding, nation B cannot afford to acquire new systems. This is implemented with a policy switch in the race submodel in section 3.

Second, with the quality strategy, hegemon nation A targets the quality of rising nation B's autonomous capabilities. It does so as soon as the performance of nation B's systems is perceived to be on the same level as nation A's. For this, on the one hand, nation A restricts nation B's access to highly skilled personnel. Without skilled personnel, nation B will not be able to achieve the level of autonomy that it desires for its autonomous capabilities. On the other hand, nation A also restricts nation B's access to the critical components that it needs for building high quality autonomous capabilities. Thus, nation A can prevent nation B from access to high quality autonomous capabilities. Without these, nation B might be able to build autonomous capabilities of a large quantity. However, by taking the quality strategy, nation A assures that the quality of these systems will remain low. This policy is implemented using two policy switches for each co-flow in the Technology Submodel in section 3.

Finally, these two strategies can be combined into one: the combined strategy. Then, if the quantity of nation B's autonomous capabilities equals nation A's, it will restrict nation B's funding. If the quality of nation B's autonomous capabilities equals nation B's, it will restrict nation B's access to skilled personnel and critical components. If both the quality and the quantity of nation B's capabilities meets those of nation A, nation A applies both strategies at the same time.

The most promising strategy in preventing arms races from happening is the combined strategy (see Figure 16). By taking this approach, the number of arms races is slightly less than 10%, as can be seen in Figure 20. This is 75% less than without any policy implemented. At the same time, the funding strategy results in 65% less arms races than when no policy is taken and the quality strategy on its own results in 30% less arms races than without any policy.

The most promising strategy for hegemon nation A to win the arms race is also the combined strategy (see Table 5), where it will win in 86% of the scenarios. In fact, the quality strategy makes hegemon nation A loose arms races more often than when nation A does not employ any strategy, and it wins in only 36% of the scenarios. When using the quantity strategy, nation A is able to win the arms races in 61% of the scenarios.

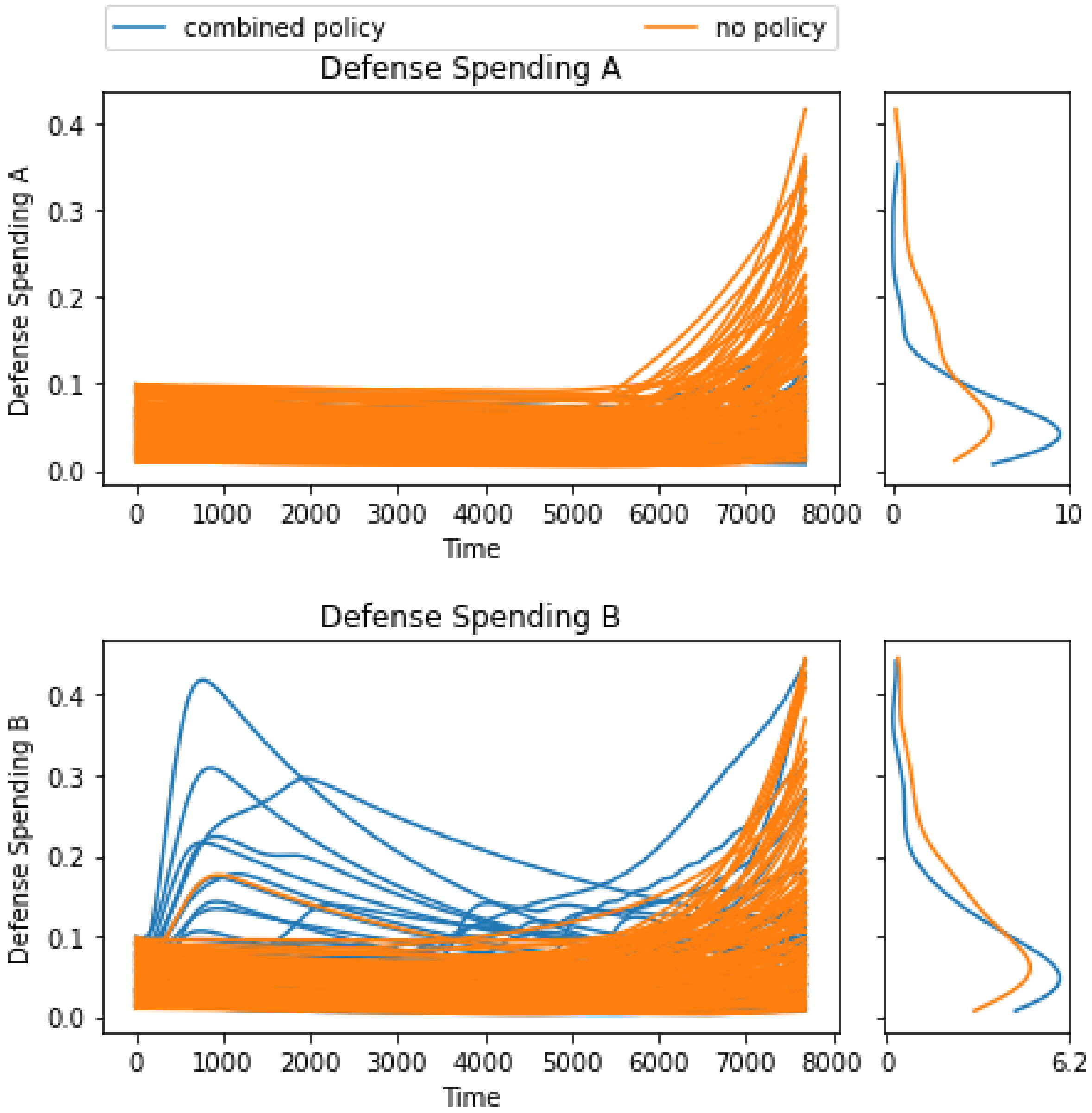


Figure 16: On the left: Defense spendings as percentage of GDP of nation A (top) and nation B (bottom) over simulation time of 30 years when nation A applies the combined strategy to target B. On the right: Kernel density estimation of the defense spendings of applying the combined policy (blue) and no policy (orange). Applying the combined policy shifts the defense spendings of both nations to lower values.

Interestingly, the best performing policy (the combined strategy) results in most races being driven unilaterally by nation B. While it overall results in the least number of arms races, this being only in 10% of the experiments, leads to only about 48% of the arms races being bilateral ones. 46% of the races under the combined policy are actually driven by nation B. Hence, while there are less arms races in general, those that do occur are due to nation B increasing investments. This is surprising, since when employing the combined policy, nation A is intentionally restricting the funding that nation B has access to. When looking at the dynamics of the defense spending of nation B over time as can be seen in Figure 16, there is in about 7% of the cases a bump in the defense spendings of nation B around time step 1000. This bump is the result of an increase in personnel cost: since nation A restricts the level of autonomy of nation B, the latter needs more operators and maintainers of its autonomous capabilities. Hence, nation B starts recruiting massively at the beginning of the simulation time and the defense spending increases as a result from training all of these newly recruited people.

When looking at one specific model run of a B driven race in which nation B wins and nation loses in Figure 17, one can see that the defense spending of nation A constantly decreases during the entire simulation time. For nation B on the other hand, there are two rather steep increases in spending. At the same time, A's defense spending falls constantly, therefore, is a race driven only by B.

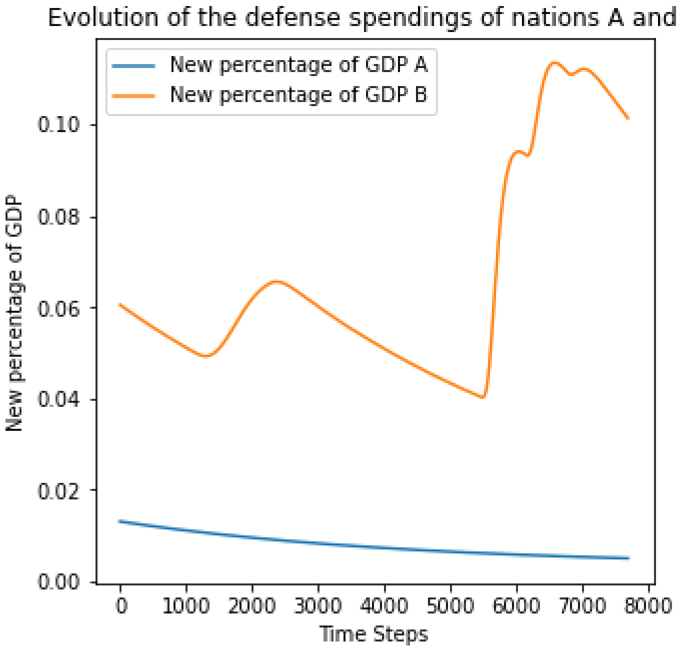


Figure 17: Defense spendings as percentage of GDP of nation A (blue) and nation B (orange) over simulation time of 30 years in one unilateral scenario where nation A applies the combined strategy

When relating this to the behavior over time of the stock of autonomous military capabilities in Figure 18, one can see that the inflection points in nation B's defense spending take place slightly after a time step in which the size of nation B's autonomous military capabilities stock overtakes nation A's size. After some delay time, the autonomous military capabilities stock of nation B eventually starts to fall, according to its falling defense spending. As soon as the quantity of nation B's stock is smaller again than nation A's, nation A will stop restricting nation B's funding, such that nation B can start building up its stock of autonomous capabilities again.

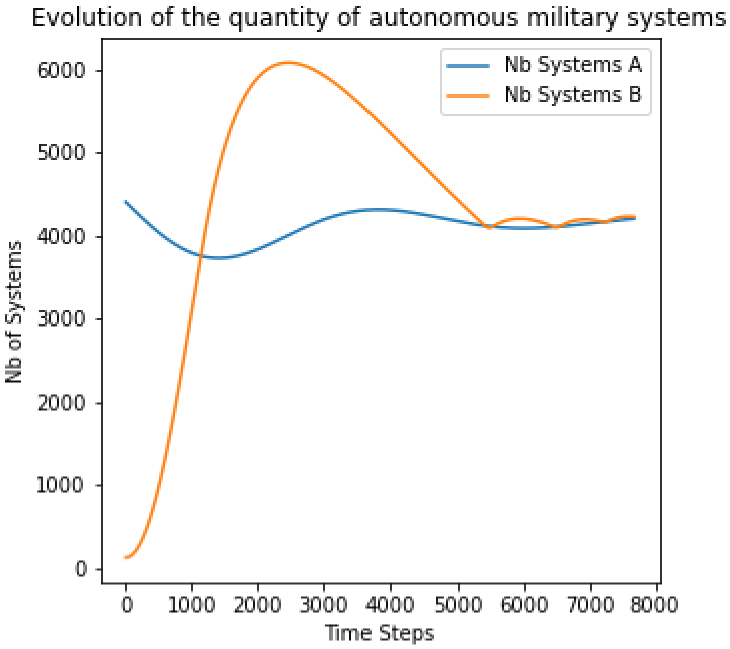


Figure 18: Behavior of autonomous capabilities of nation A (blue) and nation B (orange) over simulation time of 30 years in one unilateral scenario where nation A applies the combined strategy

As soon as it becomes greater than nation A's however, nation A will restrict B's funding again. In this specific scenario, this happens four times, and hence the four bumps in nation B's stock in autonomous military capabilities. Nation A has no desire in making up for this gap by increasing its defense spending and investing more into autonomous military capabilities: By opting for the combined strategy, it tries to assure itself an advantage in terms of quality of its autonomous military capabilities. As can be seen in Figure 19, nation A also slightly gets more autonomous systems around a point in time in which nation B overtakes it quality. After nation A restricts its potential for developing high quality systems, it acquires less systems itself. Thus, the evolution of nation A's stock of autonomous systems oscillates slightly with the evolution of B's quality. Around the end of the simulation time, nation B is able to win this specific scenario, as nation A is too slow in targeting B's quality to assure A keeps the edge.

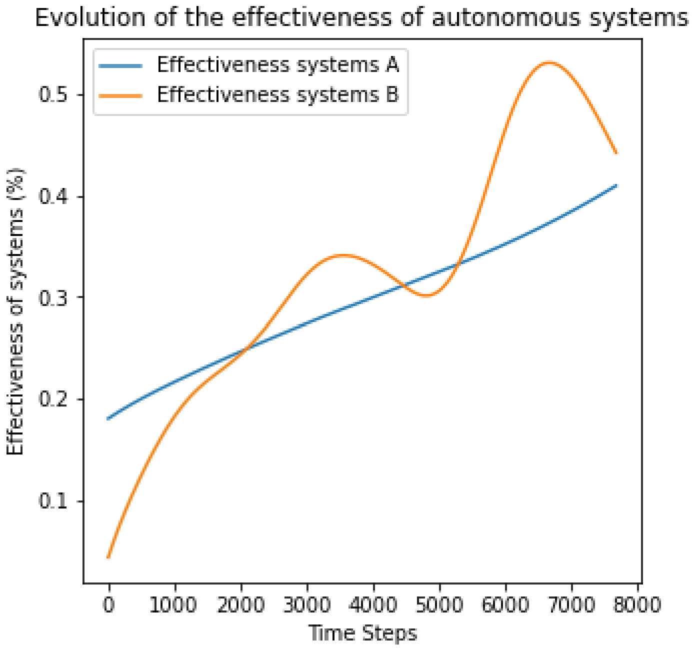


Figure 19: Behavior of the effectiveness of nation A (blue) and nation B (orange) over simulation time of 30 years in one unilateral scenario where nation A applies the combined strategy

The combined strategy is chosen as a starting point to devise a robust policy. This is the most successful strategy as it results in the least number of arms races. At the same time, employing this strategy makes nation A win the arms races most often when compared to other strategies (see Table 5). Vulnerabilities of this strategy are identified by determining which parameters characterize those scenarios in which there are still arms races happening. Therefore, as PRIM is leading to inconclusive results (see Appendix D, the relative differences of parameters values leading to arms races when applying the combined strategy versus no arms races when applying the combined strategy are computed.

Type of Race	Quality as target	Quantity as target	Combined strategy
Bilateral	16 %	55 %	100 %
A driven	31 %	50 %	100 %
B driven	15 %	51 %	71 %
No Race	54 %	68 %	85 %

Table 5: Percentage of arms races that hegemon nation A wins, depending on the different types of arms races, when only nation A targets nation B's development of autonomous capabilities.

Arms races under the best performing strategy are characterized by the following parameters: the initial stock of autonomous capabilities of nation B is lower, while A's is higher, and the general capabilities of B are lower than when there are no arms races. At the same time, A is initially spending more on autonomous systems and its GDP is growing faster than in not arms racing scenarios. Rising nation B is developing its systems faster than in not arms racing scenarios. Moreover, nation B needs less personnel for its autonomous capabilities. Finally, hegemon nation A is more reactive with a time to change its desired margin of superiority is faster than in non-arms racing scenarios.

In Table 6 the different outcomes can be compared.

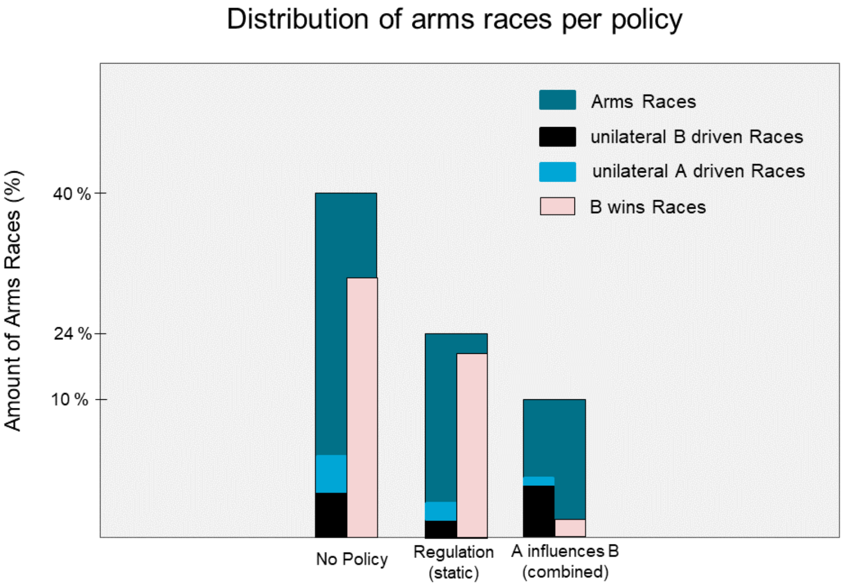


Figure 20: Distribution of different types of arms races in different model runs.

Policy	A wins Race	A wins No race	Number of Arms Races
None	24 %	50 %	40 %
Quality	18 %	54 %	35 %
Funding	54 %	68 %	14 %
Combined	87 %	85 %	10 %
Static	10 %	50 %	24 %
Midway	15 %	50 %	30 %
Adapting	10 %	48 %	32 %

Table 6: Percentage of arms races that hegemon nation A wins, depending on the different policy.

4.2.3 Nation B strikes back

So far, rising nation B had no means of influencing hegemon A. In a next step, rising nation B can strike back: it can also target nation A using the best performing, combined strategy. With this, it can restrict nation A's access to funding, to critical components and to highly skilled personnel.

When nations A and B mutually spoil each other's race for autonomy, there are no more arms races: there are no scenarios in which any nation increases its defense spending by more than 50% over 5 years. This is to be expected, as they both hinder each other from doing so. In 97% of these scenarios, nation A is able to win the competition. This corresponds to the prior observation, that nation A usually wins when there is no arms racing going on, due to the fact that it has the head start. Still, in 3% of the scenarios, nation B is able to win the competition. In these scenarios, nation A needs even more operators and more "human control" for its autonomous capabilities, while nation B needs in total less personnel for its autonomous capabilities.

5 Discussion

In the following, I discuss the implications of the model results for answering the research questions; namely about the circumstances under which arms races occur and the strategies that nations can employ to influence these. First, I contextualize the conditions under which arms races for autonomous military systems take place in the model world versus the real world. These conditions were a very weak rising nation B and a declining hegemon nation A with higher ethical standards that both are very reactive to each other's arming actions.

In addition, I discuss the strategies the nations can employ to influence arms races in the model world versus the real world. Here, nation A should keep out of arms races, since it loses them more often than not. Nation B however should engage in an arms race since then it is able to win. Consequently, nation A should try and spoil the race for autonomy for nation B. It should then do so by restricting both the quality and quantity of nation B which leads to the best results in terms of reducing the occurrence of arms races and maximizing the number of arms races in which nation A wins, even if nation B strikes back. Moreover, a third party institution that aims at avoiding arms races regardless of the objectives of nations A and B should seek to ease the tensions between nations A and B to make them less ”reactive” to each other and aim at raising the ethical standards for nation B.

Then, I include a discussion on the role of quantitative simulation modelling versus qualitative modelling in strategic defense analysis. Further, limitations of the used methodology are discussed, which includes the consultation of experts and the construction of a SD model to conduct computational experiments for scenario discovery. This finally leads to recommendations of future research on how to fill these limitations and on which new questions have been raised in the course of this research.

5.1 Conditions for Arms Races in the Model versus Real World

To begin with, arms races in the model world happen under the same conditions as real world arms races, as defined by Caspary (1967): (1) perceived military insecurity, (2) the arming being a burden on economy and (3) perceived grievances. The second condition is necessarily given by the definition to identity an arms race from (Diehl, 1985). The first and the third condition of arms races are also met by the factors that were identified to lead to arms races: Rising nation B has very low initial capabilities and as a result, perceives military insecurity due to its lower military power. Plus, arms races are characterized by relatively short times to change their desired margin of superiority for both nations as well as a higher bias in estimating the opposing autonomous military capabilities by nation B. The short times relate to short response times to each other, meaning that the two nations are very quick to react to their opponent's actions. These conditions could be seen as the nations having a high sense of urgency to build autonomous capabilities and can be interpreted as grievances between the two nations.

Moreover, arms races in the model world generally all start the same: There mostly is a more or less big gap in the quantity of autonomous military capabilities between hegemon nation A and rising nation B, with nation A having the advantage as B is generally inferior in terms of military power. In these cases, nation B increases its defense spending to grow its own autonomous military capabilities. This is based on the assumption that nation B does not consider using any other means for addressing the growing threat that nation A poses to nation B than responding by ramping up on its own autonomous military capabilities. In real world as opposed to the world of my model, the political leadership in nations A and B might be talking about their issues via diplomatic channels. Thus, nation B might in the real world, first address its growing concern about nation A's activities towards nation A before getting involved in an arms race.

Arms races then really take off, when rising nation B gets closer or even an advantage to hegemon nation A in terms of quality of its systems. On these occasions, nation A starts heavily increasing its own capabilities. This leads to the exponential growth of the nations' defense spendings starting at a given time towards the end of the simulation time. At this moment, nation B has overtaken to nation A in the quality of the systems. In some cases, where nation B only comes close to nation A and A can keep to edge, nation A exponentially increases its defense spending but then calms down after some time which leads to defense spendings reaching a plateau and an overall behavior of an S-shaped growth. Again, any other means of addressing this perceived threat are not considered. Depending on the quality of their systems, nation A and nation B mutually push each other towards increasing their autonomous military capabilities. Moreover, rising (and less ethical) nation B needs less personnel for its autonomous capabilities. Thus, it can use most of its budget to spend on the acquisition of autonomous systems as opposed to spending it on the (training) costs of personnel. Therefore, as nation A has higher ethical standards that require more people for its autonomous systems, B can get more quantity than A for the same amount of funding.

In addition, the concept of the balance of power in international relations also applies (Wohlforth, 2014): In the most extreme arms races, rising nation B is overall extremely weak compared to nation A. According to this theory, some stability and peace is assured when there is a relative military power parity between two opposing nations. If, however, the power differential becomes too high, the risk for escalation of the tensions between two nations increases as well and consequent arms races can occur.

The concept of power transition from international relations theories applies as well (Wohlforth, 2014): When a nation's power increases or decreases, the risk of escalation of tensions is greater as the nation seeks to either assure its further ascent or prevent its further decline. This is reflected with the arms races that the model produces: While nation B is weaker than nation A, it is certainly rising as can be assumed from it being able to produce high quality technology. At the same time, while nation A is stronger, it is to some extent declining: In the extreme arms races, its GDP growth is a lot smaller than in non-racing scenarios which can be an indication for its economy slowing down and with this its power.

Finally, two opposing nations are also at (lower) risk of the escalation of tensions when there are unilateral arms races as opposed to bilateral arms races (Boswinkel and Sweijs, 2022). Most of the races that occur in the model world are bilateral ones. Some however, are driven by only one nation. The respective opposing then does not get involved. Interestingly, the circumstances for unilateral arms races are like those for the extreme races: Rising nation B gets triggered to start building up autonomous capabilities, by nation A initially spending a lot on autonomous systems, while B initially spends very little. Therefore, there is an initial gap between the two nations that the nation lacking behind desires to fill. Then, there are two possible reactions of nation A to nation B's arms buildup: Either it does not get involved at all, leading to the unilateral arms races, or A does build up as well and it does so massively. The additional condition for the latter scenario is the high quality of the systems. If B is weaker and at the same time not able to produce any high quality and highly autonomous systems, nation A does not care as it does not fear to lose a qualitative advantage (and vice versa), leading to a unilateral arms race.

5.2 Strategies to influence Arms Races in the Model versus Real World

Influencing arms races can be done from different perspectives with different objectives. I assume that nations A and B aim at influencing the arms races in order to win them or avoid them, while a third party such as an intergovernmental organization might be interested in avoiding all arms races whatsoever. I make the distinction between strategies that (bilaterally) regulate the development of autonomous military capabilities and strategies that aim at targeting the respective opposing nation to spoil its race for autonomy. Furthermore, I make a difference between strategies aiming at influencing qualitative versus quantitative arms race (Intriligator and Brito, 1984).

5.2.1 Regulating the Race for Autonomy

To begin with, regulating the level of autonomy to a level of 5 according to Sheridan and Verplank (1978)'s measuring scale as was attempted in subsection 4.2 is not desirable for hegemon nation A in the world of the model. When putting a cap on the level of autonomy that a nation's autonomous capabilities are allowed to achieve, the number of arms races that hegemon nation A loses actually increases, no matter at what instance during the simulation time this regulation is taking place. In fact, the worst futures for hegemon nation A happen when the regulation is only being applied after one of the nations exceeds a level of autonomy of 5. This corresponds to a system that can suggest a plan of actions which it needs its human to approve and can be translated to the concept of ”meaningful human control” (Ekelhof, 2018). Thus, when nations are not allowed to further increase the possible level of autonomy of their systems, hegemon nation A loses the race for autonomy more often than not.

While this finding seems counter intuitive at first, it need to be considered that these regulations only restrict one of the dimensions of arms races; namely the qualitative one while the quantitative remains. When the quality of a nation's autonomous capabilities is greater than that of its opponent, it is less inclined to engage in an arms race and to increase the size of its autonomous capabilities and thus its funding, as its systems are of better quality anyways. Hence, of those scenarios, where a nation is technologically superior to its opponent, there are actually less arms races. If, however, this option is taken away, the two nations can only compete in numbers. Hence, it is counterproductive to regulate the level of autonomy: A nation then only has the option of quantity in the competition with its opponent. Then, it can be argued that the nations should mutually agree on a quantity of autonomous capabilities. However, this kind of nonproliferation agreement is unlikely to work in reality: the application of autonomous systems is too broad and due to their dual use nature, any off-the-shelf product is very quickly adopted for military purposes as can be seen in the war in Ukraine (Kramer, 2022). In fact, it remains questionable whether restricting the level of autonomy of military capabilities in itself is actually feasible. Nuclear weapons can be counted, while autonomy cannot. During the Cold War, treaties such as the Intermediate-Range Nuclear Forces Treaty were signed by the USA and the USSR to ban all intermediate range ballistic missiles and thus deescalate the arms race for nuclear weapons (Kühn and Péczeli, 2017). Here however, it will be hard to control whether nations keep their word and restrict their capabilities to the agreed amount of ”meaningful human control. Again, the dual use nature of autonomous robots poses an additional challenge, if civilian products are allowed to have a higher degree of autonomy than their military counterparts. Hence it can be argued that in contrary to Meadows (2008), disarming during the race of autonomy might not work.

In the model, regulating the level of autonomy also has an interesting side effect, due to the way in which military insecurity is determined: Next to the number of an opposing nation's autonomous capabilities, a nation also considers the performance and effectiveness of these systems. Hence, when the opposing nation has a lot of autonomous capabilities, but these have a lower performance (as a result form a lower level of autonomy), the perceived military insecurity is less and there should be less propensity to engage in arms racing. However, at the same time, due to a lower level of autonomy, a nation needs to recruit more personnel to operate its growing size of autonomous capabilities. Some of the additional arms racing in the regulated scenarios are consequently not only a race against each other's arming, but to some extent also a race against each other's recruiting. To some extent, this can be observed in real-world deterrence strategies, where countries having grievances with their neighbors have mandatory military conscription (see Taiwan, South Korea, or Finland).

5.2.2 Spoiling the Race for Autonomy

In order to win the competition for autonomous military systems, nation A should keep out of an arms race: With its head start in the technological development of high quality systems, it will win the competition more often in non-racing scenarios than in racing ones. Nation B on the contrary should engage in arms races in order to win the competition for the development of autonomous systems. This in turn suggests, that nation A should spoil the race for nation B (and vice versa, since nation B will not just sit back and watch nation A targets it but strike back).

I introduce three different strategies in subsection 2.4, these being (1) the funding strategy aimed at quantitative arms races, (2) the quality for addressing the qualitative arms races and (3) the combined strategy. While the quality strategy results in a similar dilemma as the regulations, the funding as well as the combined strategy can reduce the number of arms races that are happening. First, with the funding strategy, A can restrict nation B's access to funding by focusing on fake news campaigns that undermine rising nation B's ethical integrity. Nation A could spread news that nation B's military is developing ”killer robots” that will randomly be killing civilians. Nation A then must be successful in shaping the public opinion in nation B. A real world example of such social media campaigns are the Russian troll factories meddling in the 2016 elections of the United States MacFarquhar (2018). The government of nation B needs to react to this by restricting the access to funding for its military. Otherwise, third countries could refuse to export their technology to nation B and eventually, its access to autonomous capabilities is restricted anyways.

Second, for implementing the quality strategy, hegemon nation A can limit rising nation B's access to high quality autonomous capabilities by restricting its access to critical components that nation B needs to build high quality autonomous capabilities. These include technologies such as semiconductors that the autonomous capabilities need for computational power and sensors that they need for operating autonomously. This can be done via sanctions and trade embargoes and thus forbidding third countries to exporting these components to nation B. The effectiveness of export sanctions of said components is then debatable, as can be observed today with Dutch chips ending up in Iranian UAS used by Russia for its war in Ukraine (NOS, 2022). Nation A can also limit nation B's access to highly skilled personnel that it needs to build high quality systems. The most brutal and least ethical way of doing so would be for nation A to take out the lead scientists of nation B, similarly to how Israel is targeting Iran's nuclear scientists (with autonomous systems) (Bergman and Fassihi, 2021). Another way of doing so would be to lure them away and to assure them work in hegemon nation A at higher wages and better working conditions, such as the US did with German scientists to build their nuclear capabilities after the second world war (Golinkin, 2021). Then, it needs to be assured that rising nation B's workforce is not too patriotic and is willing to defect from its motherland.

The third strategy leads to the least number of arms races, namely when both nations combine the other two strategies and spoil each other's race for autonomy both in quantity and in quality. Nation A is then able to win most of the racing as well as non-racing scenarios, even if nation B strikes back. This combined strategy does still have several vulnerabilities however, and still leads to arms races with an even stronger hegemon nation A and an even weaker rising nation B. Here, the latter, with placing lower importance on having humans in control of is autonomous systems is still able to produce higher levels of quality and the related autonomy than nation A. While nation A is, after all, able to win most of the times, nation B still wins in those scenarios where nation A needs "too much" human control for its autonomous systems. Plus, nations A and B are highly reactive to each other's arming actions.

It needs to be noted that, for all these operations both nations needs the funding and the power (or capabilities) of doing so. To some extent, this funding will increase the two nations' defense spending, as they can be considered to some extent as part of military expenditure and as part of the arms racing. When doing so, the number of occurring arms races and the conditions under which they do might remain unchanged. Only the distribution of arms races might be affected. Those arms races that are now unilateral ones and driven by nation B might become bilateral ones, as nation A is then considered to increasing its defense spending to undermine nation B. Also, both nations are now always successful in their attempts of targeting each other. The cost of failing in any of these operations are high however: if it comes to light that for instance nation A meddles into nation B's affairs, nation A might lose its face in the international community and can face consequences, while the tensions between nations A and B most certainly worsen.

Furthermore, a third party aiming at avoiding arms races should aim at easing the tensions between nations A and B such that they are less reactive to each other. These more or less short "reaction times" come up in most arms racing scenario, no matter what strategy is employed. If the nations take more time to consider before also ramping up on autonomous military systems as a reaction to the opposing side's, more arms races can be avoided. Moreover, this third party could aim at raising nation B's ethical standards and requirements for human control of its autonomous capabilities.

Finally, it can be debated whether nation A should adapt its ethical standards and allow for less human control of its autonomous systems: nation B wins arms races under the condition that nation A requires a lot of personnel for the operation of its autonomous capabilities. Hence, it can be argued that there might be some circumstances in which it could be advisable for nation A to lower its ethical standards and allow for less human control of its autonomous capabilities if this means that it will always win a race towards autonomy against an opposing nation with very little ethical standards.

5.3 Role of Simulation Modelling in Strategic Anticipation

At TNO's department for military operations and more specifically the project team for strategic anticipation, most modelling is of a qualitative nature. Strategic anticipation includes trend analysis and driving forces identification. For this, the project team at TNO uses qualitative tools to determine the causalities between trends and then assesses their impact and their uncertainty. They generate hypotheses about the relations between trends and rank these in order to group them depending on their assessed impact and uncertainty. Based on relevant trends, they design future worlds and scenarios, using qualitative methods such as the Shell-method or the cone of plausibility.

With this thesis, I explore the potential added value of simulation modeling for strategic anticipation. The model in this research also includes the causalities between different indicators. The main difference is that here, I can simulate these causal relations over time, and I can conduct computational experiments. By varying the parameter values of these indicators, I can compute and generate different scenarios and analyze the resulting different dynamics over time. Still, the quantification of the simulation model was challenging: The numbers that I use in the model to compute arms races are based on assumptions and estimations and as a result, are highly uncertain (e.g., the initial number of autonomous capabilities a nation possesses, or the number of people needed to operate one system). Thus, the numbers that the model computes need not be taken for ”real” values. Some variables are, due to their nature, hard to quantify. The tools offered by the EMA workbench allow to still apply quantitative simulation modelling to problems where numbers are hard to find. At the same time, this then bears the risk of the researcher not looking hard enough for the right numbers as these can be put off as uncertainties. Also, some concepts that I need in the model are dimensionless, such as the level of autonomy or technological sophistication which makes quantification challenging again.

Due to the nature of open exploration that I used to generate scenarios with the model, I chose a uniform distribution for sampling over the uncertainty space with the EMA workbench. Therefore, it cannot be distinguished between probable and less probable worlds. This is in contrast with how scenarios are often generated at the strategic anticipation team at TNO, where a cone of plausibility is used for the design of scenarios and only a couple of scenarios consist in a future world that is as good as impossible. At the same time, however, open exploration allows to test the entire space of possible worlds and all the behaviors that the model is able of generating.

Regardless of these challenges, I found that having numbers helped me to contextualize the scenarios of interest and to better understand the specific conditions under which arms races arise. Again, the exact numbers of the parameters that represent these conditions are no predictions. But still, it makes two different scenarios and two different future worlds comparable. Hence, the added value of simulation modelling is in my opinion the quantified causal connection of different indicators of trends and the simulation of these causal connections over time.

5.4 Limitations

In this section, I discuss the limitations of this research. These limitations consider the methodology for constructing the model and conducting computational experiments as well as my personal bias as a researcher.

5.4.1 Methodological Limitations

The world of the model is limited in its scope and hence in reflecting the real world in its complexity. To begin with, in the world of the model, nations perceive their military insecurity based on the size of personnel of the other nation, the other nation’s size of general military capabilities and the size of its autonomous military capabilities. Nations then desire to increase their autonomous military capabilities based on their perception of their opponent’s autonomous military capabilities stock. In the real world however, autonomous military capabilities are used for a plethora of reasons. It then depends on this specific application in a specific use case whether autonomous military capabilities are actually used to attack another nation or whether they are used by a nation to defend its interests. Hence, it cannot be assumed that all autonomous systems are necessarily a threat as can be assumed with nuclear weapons of which the application might result in mutual destruction. This is where the offensive versus the defensive applications comes into play that is discussed in some models of international relations theory. In my model, nations do not make a difference between their opponent's offensive or defense autonomous military capabilities, mainly since the technology itself is neither only offensive or defense but its application is. Nevertheless, it can be argued that autonomous military capabilities that are used in a defensive way to a certain extent also decrease the perceived security of the other nation (Wohlforth, 2014). Hence, it is deemed to be a valid assumption that also defensive capabilities are perceived as threat.

Furthermore, in the model world, one aspect of military insecurity is given by the size of military personnel. Recruiting in the model is done based only on the need for personnel due the increase in the size of autonomous capabilities. However, in a real life great power tension and related security dilemma, nations will be recruiting also based on other factors than only needing people to operate and maintain the military's capabilities. Regardless of this, the model structure can be seen as considering the lack of personnel and the problem of militaries to recruit personnel to fill their vacancies. Hence, the assumption that recruiting is driven only by the need to operate and maintain autonomous military capabilities is reflecting this. After all, the goal of the Dutch armed forces in relation to autonomous systems is to achieve three times as much with three times less people or, in other words, to be nine times more effective.

Adding to this, grievances are represented by perception factors such as the bias in estimating the opposing nation's size of autonomous military capabilities. These parameters are constant in the model, while in reality there might be some dynamics underlying these factors. In real life, this bias does not remain the same and evolves over time, as could be observed during the Cold War and the proclaimed "missile gap" by J.F. Kennedy where the USA presumably lacked behind the USSR in their ballistic missile arsenal (Preble, 2003). When the perceived quantity or quality of opposing technology increases, so could the bias. This would result in a "dynamic bias" which is not captured in the model structure. The same holds true for the technology weighting factors which are now constants in the model. However, based on perceiving the opponent's quality of AI increasing, a nation might place a higher weight on AI as well.

Next to military insecurity and grievances, the arming being a burden on the economy is another condition for an arms race. Those arms races that result in extreme defense spendings are actually happening when the GDP growth rate of nation A is smaller than on average. In these, the defense spending of nation A is growing faster than the GDP. However, economy is not explicitly included in the model. On the one hand, it can be assumed, that a defense spending of more than 10% is a burden on a nation's economy that is hard to sustain. On the other hand, when a nation has a domestic defense industry that would benefit from this extensive defense expenditure, economy would benefit from it and consequently grow, leading to a growth of GDP. The burden on economy might also be alleviated by ongoing research and development and a further reduction in the cost of the autonomous systems. Furthermore, trade in general and with the autonomous robots as dual use technology would also greatly benefit a nation's economy and might make this defense burden more bearable. While to some extent, countries will not trade their military technology, at least not with nations they are in conflict with, they might strip these systems of their critical and sensitive parts and trade those civilian versions of their technologies.

Moreover, the effectiveness of the autonomous systems increases with increasing level of autonomy. This is based on the assumption of an ordinal scale. One gains the same advantage in performance and effectiveness when increasing the level of autonomy from 1 to 2 as when increasing it from 9 to 10. Whether this reflects reality is debatable. To illustrate, according to Sheridan and Verplank (1978), a level 1 system provides no assistance whatsoever, while a level 2 system offers a complete set of action alternative to its human. It can be said this is a considerable benefit. A level 9 system automatically executes its own designed action plan and afterwards, sometimes informs its human, if necessary. A level 10 system on the other hand decided everything and in fact ignores its human. Here, the gain advantage is not as obvious anymore. Thus, the scale for measuring autonomy and the related effectiveness of the systems could impact the occurrence of arms races produced by the model. Also, in the model world, the two nations make to some extent a distinction between autonomy and technological sophistication. In reality, these two are not independent of each other and a trade-off can negatively affect the performance and effectiveness of the autonomous systems. To illustrate, without a great enough amount of technological sophistication of a system, it is difficult to introduce autonomy and the necessary components of making autonomy possible. For an autonomous system to operate autonomously, a great number of sensors is needed which produce a vast amount of data that the system's processing unit needs to process to provide accurate feedback to its actuators and to take decisions. In general, an accurate estimation of the effectiveness of autonomous systems also depends greatly on the exact application of these systems in their operational context.

In the model world, the strategies for a nation to confront its opponent address targeting its development of autonomous capabilities. The selection of a target is modeled solely based on the criticality of the potential target for the opponent. This means that when the opponent chooses the quantity strategy for the race towards of autonomy (i.e., piling up autonomous systems), then the funding is more critical for the opponent than for instance its skilled personnel. If, however, the opponent opts for the quality strategy (and thus aims at developing high quality systems), its personnel is more critical than the funding. In real world, the targeting process is way more complicated (Ekelhof, 2018). In real world, a nation would also have many different choices to address its concerns with another nation, for instance via diplomatic channels. Also, strategies for reducing the tensions in real world could include trade policies, not necessarily with their military systems and technology but with other goods. This way, balancing mechanisms that aim at building trust would be introduced between the two nations. At the same time, a third nation or a third party would in real world try to help two conflicting nations to resolve their issues.

Finally, arms races are not necessarily a bad thing: Research shows that arms races are not usually leading to wars (Boswinkel and Sweijs, 2022) and that arms races can lead to peace as often as disarmament can in fact lead to instability and conflict (Intriligator and Brito, 1984). After all, while nations are focused on acquiring technology, they do not focus on fighting a war (Gray, 1971). Then, it remains debatable whether it is desirable to have nations piling up on autonomous systems that some refer to as ”killer robots” (Horowitz, 2016): Autonomous systems in themselves have uncertain effects on international stability, as pointed out in Altmann and Sauer (2017). This is due to three conditions, namely (1) the fact that they are available at cheaper prices, (2) can blur the line between defensive and offensive purposes (which can increase the perceived military insecurity of a nation) (3) and have less life at stake. Especially when there is a high sense of urgency to develop autonomous military capabilities, nations might be willing to disregard their own ethical standards while developing autonomous systems that then do not comply by international humanitarian law.

5.4.2 Personal Bias

I built the model for this research based on the available literature and experts that agreed on contributing their knowledge. Plus, I interpret the results from the model runs and draw conclusions on the latter, based on by my personal reflections. As a result, there are several instances where these conclusions are not entirely objective but influenced by my subjective bias.

To begin with, while I was reviewing literature from differing backgrounds, the experts I consulted all share similarities which can lead to a certain degree of bias in the results. First, the experts I interviewed are all based in The Netherlands and hence all have a ”Western” perspective. Here, I would have liked to reach out to people from different corners of the world, if time had allowed for it. Now, the two nations that I include in my model have the same assumptions underlying their choices, namely mine Western ones and the ones from my Western experts. At the same time, a nation with a different culture might interpret winning an arms race very differently and assess the winning strategies in a different way.

Then, my experts are all affiliated to the defense domain. While I had also reached out to people with different background, I had not been able to arrange interviews with for instance politicians or stakeholders from peace organizations. The latter might have a very different view on the development of autonomous military capabilities and the perceived threat that they pose. However, since in this thesis I aim at modelling how two nations and their militaries are engaged in an arms race, it seems to be a valid choice to mostly take into account the point of view of military experts.

Moreover, the majority of my interviewees are male. While most topics I discussed with my experts are not necessarily gender specific (e.g., technology), the perception of threat and security from a male point of view could potentially differ from a female point of view. Here again, in this thesis I am modelling two nations and their militaries and since these institutions are still run mostly by men, this gender bias seems only to reflect reality.

Finally, the conclusions I draw from my model results depend on my assumptions and on the point of view that I take. When evaluating the strategies that nations can employ in the model to influence arms races, I interpret a good strategy as one where hegemon nation A wins. I could also have taken the other stance of rising nation B. However, as I interpret hegemon nation A as the more ”Western” one and I constructed the model with my ”Western” bias, I found it more suitable to also take the point of view of nation A.

5.5 Future Research

In this section, I identify some recommendations for future research. Some of these stem from the discussion about the conditions under which arms races arise and how these can be influenced. Others are a result from the identified limitations as introduced above.

First, to reiterate, de-escalation measures aiming at restricting the quality and the quantity of autonomous military systems might not work. Rather, de-escalation attempts aiming at easing the tensions between nations A and B might be more fruitful. A different attempt regarding regulations could be to restrict the application of autonomous military capabilities to only defensive purposes. That then requires the trustworthiness of both nations to keep to these restrictions. This I could not test with the model, as there is no difference between defensive and offensive technology. To evaluate to what extent such a regulation would be successful in reducing arms races, a distinction would need to be made between all the different applications of autonomous systems and their resulting effect on the perception of threat by the respective opposing side.

Next, to really get a grasp of the effectiveness of autonomous military systems, further submodels of different types and their subsystems (ground vehicles versus aerial systems, different sizes and classes, battery, fuselage, payloads etc.) could be constructed. Then, the technological developments that influence their performance for their intended applications could be modelled. Here, one could think of modelling the development of battery technology and aerodynamics of the fuselage such that for instance the implications for the weight and loitering time of a drone and hence the increase in perceived threat of an opponent can be estimated. To some extent, this would then capture not only the race for autonomy, but several races for several technologies. Consequently, the model could also be used to give advice on which technological development a nation should focus to get or keep a technological advantage over an opponent. For further determining the effectiveness of autonomous capabilities for a specific application, an attrition model could be constructed that aims at modelling the performance of the latter during a conflict, such as Jia et al. (2019) for the effectiveness of swarming of UAS.

Furthermore, in the real world, decisions to target an opponent are taken differently than in the rather simple way I have introduced it in the model world. Thus, to evaluate the strategies for influencing the opponent's development of autonomous systems more realistically, the targeting process should be modelling in more detail. Also, the means a nation has for actually applying these strategies need to be included. Now in the model, I also assumed that a nation is successful in its operations to target the opposing nation and that all these eventually have the same effect. Also, a nation cannot counter its opponent's attempts in spoiling its race. The model could be extended by a method to evaluate the most effective target in order win the arms race or to prevent the opposing nation from winning the arms race. One such method is the computation of the CARVER matrix (Greaver et al., 2018), which allows the quantification of the target selection.

Finally, next to benefiting a nation's economy, trade can also influence the tensions between two nations. While a mutual trade dependence can have stabilizing effects, a unilateral dependence can make tensions worse (Boswinkel and Sweijs, 2022). Then, the nation that is less dependent has a greater potential of harming its opponent. When considering trade with autonomous systems in the model world, it can be argued that in those scenarios in which arms races are happening, nation B has an advantage in the development of AI and autonomy over nation A. Thus, nation A might even be dependent on nation B for the autonomy of its systems. This could further worsen their tensions and have a destabilizing effect. Thus, extending the model with implementing trade can give insight into this dynamic.

6 Conclusion

With the current world order under ongoing tensions, two nations in a power transition could find themselves in a security dilemma and potentially in an arms race for new military technology, namely autonomous systems. With ongoing research and development, these systems become more and more autonomous with less human control needed. As autonomous military capabilities can be used for a plethora of applications, from target acquisition to swarming, many militaries are already experimenting with concepts to employ them.

In this thesis I investigate the considerations two nations and their militaries make when deciding upon the development of autonomous capabilities. For this, I consulted various experts in defense research. As a result, I formed a conceptualized model in section 2. Upon this, I constructed an SD model (see section 3) which I then use to determine under which conditions arms races between two fictitious nations arise. Further on, I implement strategies to influence arms races and investigate their performance. The results of these analyses and their implications can be found in section 4 and section 5.

To put it in a nutshell, the conditions for arms races I identify are the following: arms races start with a superior nation having an edge on both quantity and quality over an inferior nation. Arms races take off as soon as the inferior nation with lower ethical standards suddenly gets an edge in developing high quality autonomous systems. To the superior nation, the number of military systems that this inferior nation possesses, matters less. When the superior nation loses its advantage in quality however, and even though it still has an advantage in quantity, it starts piling up autonomous military systems. To this, the inferior nation then responds by doing the same. Thus, the model identifies arms races as situations that can be described as great power transitions with a balance of power between two nations that is not even. Hence, international relations theories such as realism do apply in the model (Wohlforth, 2014).

The best way of influencing an arms race depends on the objectives in mind. If the goal is to win, the superior nation should avoid getting into an arms race in the first place: the model results suggest that once in an arms race, this superior nation will lose the arms race more often than not. The inferior nation on the contrary should engage in an arms race with the superior nation since it wins most of the times. In fact, the inferior nation should encourage the application of regulations, since with a cap on autonomy, it wins even more frequently. The superior nation should aim at spoiling the arms race for the inferior nation in quantity and quality, as in these cases, even if the inferior nation strikes back, it is able to win the arms races in most scenarios. Controversially, the inferior nation is still able to win in some cases, namely those where nation A requires ”too much” human control of its autonomous military capabilities. If the goal is to reduce the occurrence of arms races, both nations should spoil each other’s race for autonomy as in this case, assuming the two nations are successful in doing so, there will be no more arms races.

Contrary to what Meadows (2008) proposes as the only way of getting out of an arms race, the model results suggest that bilateral disarmament might not be desirable for the superior nation, assuming this is actually feasible given the dual use nature of autonomous systems: When limiting the allowed level of autonomy of military capabilities, a nation will choose quantity over quality. When limiting quantity, a nation will choose quality. At the same time, the superior nation loses more often than not, when the level of autonomy is restricted. Another option remains, that I could not test with the model, which is the restriction of autonomous military capabilities to only defense application.

The winning strategy, namely targeting the quality and quantity of the race for autonomy, still leads to some arms races: It does not work in situations where the superior nation is highly reactive to the inferior nation's arming actions and when the inferior nation places very little importance on having humans in control of its autonomous systems. Consequently, a third party such as the United Nation could step in to facilitate a relaxation of the tensions between the two nations and ensure that the superior nation is less reactive. At the same time, this third party could encourage the inferior nation to increase its ethical standards.

To conclude, the model in which I formulate these advises is based on several assumptions. These are necessary to maintain a feasible scope. The quantitative nature of the simulation model, in contrast to a qualitative one, as often used in strategic anticipation, enables a different line of thinking: The concepts in the model world need to be measurable and hence, the results will also be measurable. This then enables the use of different tools that can be employed to analyze the scenarios generated with the model. Recommendations for future research include further improving the model by introducing balancing effects such as trade, taking a closer look at the actual effectiveness of autonomous military systems given their application, and a more realistic implementation of the targeting process for influencing an opposing nation's race for autonomy.

References

J. Altmann and F. Sauer. Autonomous weapon systems and strategic stability. *Survival*, 59(5): 117–142, 2017. doi: https://doi.org/10.1080/00396338.2017.1375263.

R. Bergman and F. Fassihi. The scientist and the a.i.-assisted, remote-control killing machine, 2021. URL https://www.nytimes.com/2021/09/18/world/middleeast/iran-nuclear-fakhrizadeh-assassination-israel.html.

L. Boswinkel and T. Sweijs. Wars to come, europeans to act: A multimethod foresight study into europe's military future, 2022. URL https://hcss.nl/wp-content/uploads/2022/10/Wars-to-come-Europeans-to-act-full-report-HCSS-2022-V2.pdf.

A. Calcara, A. Gilli, M. Gilli, R. Marchetti, and I. Zaccagnini. Why drones have not revolutionized war: The enduring hider-finder competition in air warfare. *International Security*, 46(4): 130–171, 2022. doi: https://doi.org/10.1162/isec_a_00431.

W. R. Caspary. Richardson's model of arms races: description, critique, and an alternative model. *International Studies Quarterly*, 11(1):63–88, 1967. URL https://www.jstor.org/stable/3013990.

K. Charmaz. *Constructing grounded theory*. sage, 2014. ISBN 9780857029140.

China.org.cn. President xi's remarks at the cica shanghai summit, 2014. URL http://www.china.org.cn/chinese/2014-06/03/content_32561159_2.htm.

chinascope. Silent contest, 2014. URL http://chinascope.org/archives/6447.

P. Collier and A. Hoeffler. Military expenditure: Threats, aid, and arms races. *Aid, and Arms Races (November 2002)*, 2002.

P. F. Diehl. Arms races to war: Testing some empirical linkages. *The Sociological Quarterly*, 26 (3):331–349, 1985.

K. M. Eisenhardt and M. E. Graebner. Theory building from cases: Opportunities and challenges. *Academy of management journal*, 50(1):25–32, 2007.

M. A. Ekelhof. Lifting the fog of targeting. *Naval War College Review*, 71(3):61–95, 2018.

M. Eslami. Iran's drone supply to russia and changing dynamics of the ukraine war. *Journal for Peace and Nuclear Disarmament*, 5(2):507–518, 2022.

P. D. Fajgelbaum and A. K. Khandelwal. The economic impacts of the us–china trade war. *Annual Review of Economics*, 14:205–228, 2022. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-051420-110410.

D. Fiott. Europe and the pentagon's third offset strategy. *The RUSI journal*, 161(1):26–31, 2016. doi: https://www.doi.org/10.1080/03071847.2016.1152118.

U. Flick. *An introduction to qualitative research*. sage, 2022.

L. Freedman. Introduction—the evolution of deterrence strategy and research. In *NL ARMS Netherlands Annual Review of Military Studies 2020*, pages 1–10. Springer, 2021. URL https://link.springer.com/chapter/10.1007/978-94-6265-419-8_1.

D. Gettinger. Drone databook update: March 2020. *Center for the Study of the Drone at Bard College*, 2020.

L. Golinkin. Why do stanford, harvard and nasa still honor a nazi past?, 2021. URL https://www.nytimes.com/2022/12/13/opinion/stanford-harvard-nasa-nazi-scientists.html.

C. S. Gray. The arms race phenomenon. *World Politics*, 24(1):39–79, 1971. doi: https://doi.org/10.2307/2009706.

B. Greaver, L. Raabe, W. P. Fox, and R. E. Burks. Carver 2.0: integrating the analytical hierarchy process's multi-attribute decision-making weighting scheme for a center of gravity vulnerability analysis for us special operations forces. *The Journal of Defense Modeling and Simulation*, 15(1):111–120, 2018.

D. G. Groves and R. J. Lempert. A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, 17(1):73–85, 2007. doi: https://doi.org/10.1016/j.gloenvcha.2006.11.006.

M. C. Horowitz. Public opinion and the politics of the killer robots debate. *Research & Politics*, 3(1):2053168015627183, 2016.

M. C. Horowitz. When speed kills: Lethal autonomous weapon systems, deterrence and stability. *Journal of Strategic Studies*, 42(6):764–788, 2019. doi: https://doi.org/10.1080/01402390.2019.1621174.

In.Europa. State power index, 2017. URL http://index.ineuropa.pl/en/measures-of-state-power/how-we-describe-state-power/.

M. D. Intriligator and D. L. Brito. Can arms races lead to the outbreak of war? *Journal of Conflict Resolution*, 28(1):63–84, 1984.

N. Jia, Z. Yang, and K. Yang. Operational effectiveness evaluation of the swarming uavs combat system based on a system dynamics model. *IEEE Access*, 7:25209–25224, 2019.

G. Király and P. Miskolczi. Dynamics of participation: System dynamics and participation—an empirical review. *Systems Research and Behavioral Science*, 36(2):199–210, 2019.

J. Kirshner. The tragedy of offensive realism: Classical realism and the rise of china. *European journal of international relations*, 18(1):53–75, 2012. doi: https://www.doi.10.1177/1354066110373949.

A. E. Kramer. From the workshop to the war: Creative use of drones lifts ukraine, 2022. URL https://www.nytimes.com/2022/08/10/world/europe/ukraine-drones.html.

D. P. Kreutzer. A microcomputer workshop for exploring the dynamics of arms races. In *Procedings of the System Dynamics Gourp International Conference, Keystone Colorado*, 1985. URL https://proceedings.systemdynamics.org/1985/proceed/kreut463.pdf.

U. Kühn and A. Péczeli. Russia, nato, and the inf treaty. *Strategic Studies Quarterly*, 11(1):66–99, 2017.

J. H. Kwakkel. Ema workbench docs, 2022. URL https://emaworkbench.readthedocs.io/en/latest/indepth_tutorial/open-exploration.html#advanced-analysis.

J. H. Kwakkel, W. L. Auping, and E. Pruyt. Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting and Social Change*, 80(4):789–800, 2013. doi: https://doi.org/10.1016/j.techfore.2012.09.012.

R. J. Lempert, D. G. Groves, S. W. Popper, and S. C. Bankes. A general, analytic method for generating robust strategies and narrative scenarios. *Management science*, 52(4):514–528, 2006. doi: https://doi.org/10.1287/mnsc.1050.0472.

L. Lonardo. Power to the people. how open technological innovation is arming tomorrow's terrorists, 2021.

N. MacFarquhar. Inside the russian troll factory: Zombies and a breakneck pace, 2018. URL https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html.

J. McDonald. China's baidu races waymo, gm to develop self-driving cars, 2022. URL https://thediplomat.com/2022/06/chinas-baidu-races-waymo-gm-to-develop-self-driving-cars/.

A. McKay, A. Watson, and M. Karlshøj-Pedersen. *Remote Warfare: Interdisciplinary Perspectives*. E-International Relations publishing, 2021.

D. H. Meadows. *Thinking in systems: A primer*. chelsea green publishing, 2008.

J. J. Mearsheimer. China's unpeaceful rise. In *Realism Reader*, pages 464–467. Routledge, 2014. ISBN 9781315858579. URL https://www.taylorfrancis.com/chapters/edit/10.4324/9781315858579-72/china-unpeaceful-rise-john-mearsheimer.

A. H. Michel. *Unarmed and Dangerous: Lethal Applications of Non-weaponized Drones*. Center for the Study of the Drone at Bard College, 2020.

L. S. M. Michel Rademaker. Robotic and autonomous systems: From design to development and use in military operations, 2022. URL https://hcss.nl/report/robotic-autonomous-systems-from-design-development-use-in-military-operations/.

MoD, Concepts and Doctrine Centre. Human-machine teaming (jcn 1/18), 2018. URL https://www.gov.uk/government/publications/human-machine-teaming-jcn-118f.

NOS. Nederlandse chips in iraanse drones van russisch leger, 2022. URL https://nos.nl/artikel/2455866-nederlandse-chips-in-iraanse-drones-van-russisch-leger.

Pallas Athena Podcast. Drones en Helicopters - Mark Voskuijl. URL https://open.spotify.com/embed/episode/7jntp31tVPL6yzX5y5NgAY?utm_source=generator.

C. A. Preble. "who ever believed in the 'missile gap'?": John f. kennedy and the politics of national security. *Presidential Studies Quarterly*, 33(4):801–826, 2003.

P. M. Senge and J. W. Forrester. Tests for building confidence in system dynamics models. *System dynamics, TIMS studies in management sciences*, 14:209–228, 1980.

T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.

J. Sterman. Truth and beauty: validation and model testing. *Business Dynamics: Systems thinking and modeling for a complex world*, 1:845–892, 2000.

J. Sterman. System dynamics: systems thinking and modeling for a complex world. 2002. URL http://hdl.handle.net/1721.1/102741.

J. Sterman. *Business dynamics*. Irwin/McGraw-Hill c2000.., 2010.

T. Sweijs and F. Osinga. Conclusion: Insights from theory and practice. *NL ARMS*, page 503, 2021. URL https://library.oapen.org/bitstream/handle/20.500.12657/47298/9789462654198.pdf?sequence=1#page=511.

J. Teer, T. Sweijs, P. van Hooft, L. Boswinkel, J. Eijkelkamp, and J. Thompson. China's military rise. 2021. URL https://docs.clingendael.org/sites/docs/files/2021-11/Chinas%20Military%20Rise.pdf.

The Walking Soldier Podcast. 06 The Walking Soldier: Op zoek naar meerwaarde van ROBOTIC WARFARE met Lkol Martijn Hädicke. URL https://www.youtube.com/watch?v=Xl4gL-aGbtU.

M. Voskuijl. Performance analysis and design of loitering munitions: A comprehensive technical survey of recent developments. *Defence Technology*, 18(3):325–343, 2022.

M. Voskuijl, T. Dekkers, and R. Savelsberg. Flight performance analysis of the samad attack drones operated by houthi armed forces. *Science & Global Security*, 28(3):113–134, 2020.

T. Waldman. Vicarious warfare: The counterproductive consequences of modern american military practice. *Contemporary security policy*, 39(2):181–205, 2018. doi: https://www.doi.org/10.1080/13523260.2017.1393201.

W. E. Walker, V. A. Marchau, and J. H. Kwakkel. Uncertainty in the framework of policy analysis. In *Public policy analysis*, pages 215–261. Springer, 2013. URL https://link.springer.com/chapter/10.1007/978-1-4614-4602-6_9.

D. Wars. Who has armed drones?, 2021. URL https://dronewars.net/who-has-armed-drones/.

C. Welt and A. S. Bowen. Azerbaijan and armenia: The nagorno-karabakh conflict. LIBRARY OF CONGRESS WASHINGTON DC, 2021.

W. C. Wohlforth. The stability of a unipolar world. In *Realism Reader*, pages 383–395. Routledge, 2014. URL https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9781315858579-61&type=chapterpdf.

J. D. Yuan. *Asia-Pacific security: China's conditional multilateralism and great power entente*. Strategic Studies Institute, US Army War College, 2000. URL https://books.google.com/books?hl=en&lr=&id=XJHUAAakIAwC&oi=fnd&pg=PA1&dq=Asia-Pacific+security:+China%27s+conditional+multilateralism+and+great+power+entente+yuan+&ots=TDSjwoXrvk&sig=aArXW2vDfq-HB25CxHRvLHmwaLE.

A Interview Questions

Some of the knowledge that my experts shared with me goes back to answering the questions that follow, clustered into three groups.

Technological Developments of autonomous military capabilities:

1. If you had to name three technological developments that influence the effectiveness of autonomous military capabilities, which three would you choose and why? (Altmann and Sauer, 2017; Lonardo, 2021)

2. How do these developments influence the effectiveness of different types of autonomous military capabilities for different missions? (Altmann and Sauer, 2017; Lonardo, 2021)

3. What would you estimate the average lifetime of different types of autonomous military capabilities to be? Depending on the pace of technological developments, how long would you say do armies want to keep their autonomous military capabilities before investing into newer and more high-tech ones? When would you estimate can full autonomy for autonomous military capabilities be achieved?

Geopolitical Implications of autonomous military capabilities:

1. How would you say an increase of autonomy and decrease of human control influences how a nation perceives the threat and power of another nation? (McKay et al., 2021)

2. How would you say can nations deter each other from developing fully autonomous military capabilities without restricting their own technological developments? (Calcara et al., 2022)

UAS as autonomous military capabilities:

1. What would you say is the added value of UAS as military capabilities?(McKay et al., 2021)

2. What if there was no more human control and only fully autonomous drones? How would this impact the need and deployment other military capabilities? How would this impact need for personnel? (McKay et al., 2021)

3. How does capability planning work with different types of UAS? How is the number of different types of UAS to be procured estimated? What are the specifics that make one type and brand of UAS more attractive than another? Caspary (1967)

4. Would you say UAS were indeed a ”magical bullet” as some argue, leading to an ”unmanned revolution in military affairs”?

5. How does the war in Ukraine make you change or not change your opinion about UAS as effective military capability?

B Initial Causal Loop Diagram

In Figure 21, the initial causal loop diagram is shown. In the following, I explain those effects that are taken out in the main section.

The offensive-defensive balance between the nations enforces the escalation mechanism. If nation A builds up more on offensive autonomous capabilities, the threat it poses to nation B leads to B increasing its preferred margin of superiority (R1 for nation A and R6 for nation B). If nation A builds up more on defensive autonomous capabilities, the threat it poses to B increases less as A can only use these capabilities to defend itself and not to launch an offensive on B. However, ramping up on defensive capabilities still increases the perceived threat, even if to a lesser extent. This is due to the other nation consequently ramping up on offensive capabilities to evade its opponent's defensive ones.

Trade can have both a deescalating and an escalating effect. Engaging in trading relations between the nations decreases the threat posed by the nations (balancing loops B2 for nation A and B6 for nation B). It is assumed that such policies that aim at ”deterrence by entanglement”, due to the fact of having interdependent economies, could lead to more stability in the system. If, however, for instance due to decoupling of economies, the cooperation between the two nation decreases, this balancing effect vanishes as well. On the other hand, With more trade, there is also an increase of a nation's (economic) power (R3 for nation A and R5 for nation B). This is due to the leverage it has in terms of trade of the other nation. A large industrial base will further enable large scale production of autonomous systems that can be used for military purposes. Hence, the security dilemma is reinforced.

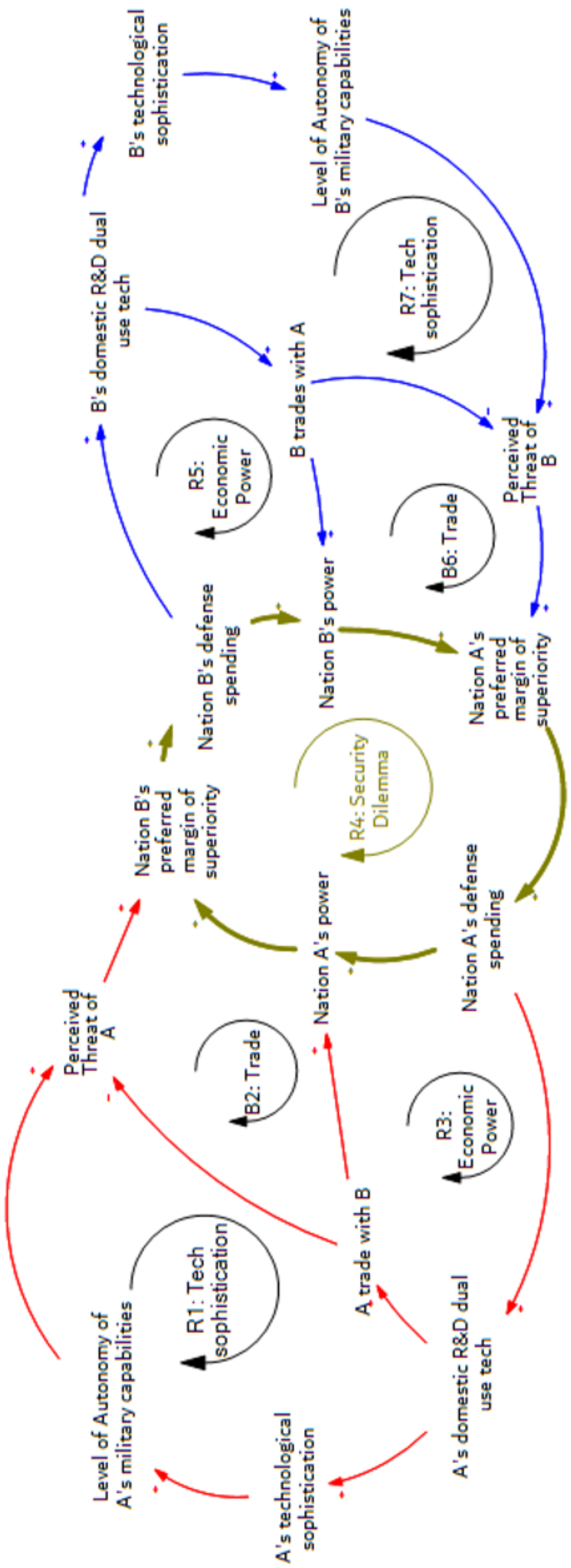

B's technological sophistication
Level of Autonomy of B's military capabilities
B's domestic R&D dual use tech
R7: Tech sophistication
B trades with A
Perceived Threat of B
R5: Economic Power
B6: Trade
Nation B's defense spending
Nation B's power
Nation A's preferred margin of superiority
R4: Security Dilemma
Nation B's preferred margin of superiority
Nation A's power
Nation A's defense spending
Perceived Threat of A
B2: Trade
R3: Economic Power
R1: Tech sophistication
A trade with B
Level of Autonomy of A's military capabilities
A's domestic R&D dual use tech
A's technological sophistication

Figure 21: Causal Loop Diagram depicting the initial dynamic hypothesis

C Additional validation

C.1 Extreme Conditions

Extreme condition tests, next to behavior-sensitivity tests, aim at exploring which behavior the model generates when the uncertain parameters are set on extremely high or extremely low variables. They are also done with the EMA workbench at the same settings as the open exploration, conducting 1000 experiments by combining the halved lowest and doubled highest possible value for each parameter. When performing experiments at extreme parameter values, about 63% of the scenarios result in extreme defense expenditures of more than 46% of GDP and 75% in arms races.

A pattern can be observed at realistic defense spendings shown in Figure 22. In this scatterplot, the final defense spendings at year 30 of the simulation time are shown. To avoid cluttering of the figure, the extreme defense spendings of higher than 46% of GDP are not shown. The reaction to each other's defense spending happens in four clusters, due to the combination of low and high parameter values: As long as the defense spending of the opposing nation does not exceed a certain threshold, the domestic defense spending remains below a certain value. As soon as it exceeds this threshold, the domestic defense spending is set at a higher value. In each cluster, high defense spendings of one nation coincide with a low defense spending of the other nation. A limited number of scenarios show a somewhat linear relation between the defense spendings of the two nations.

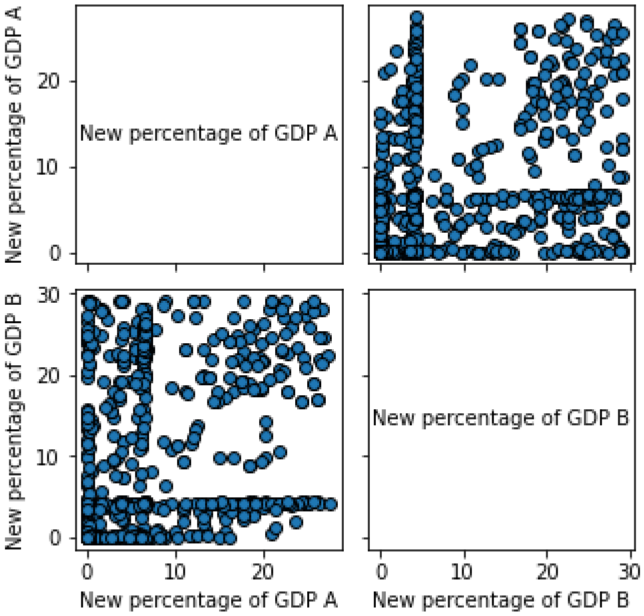


Figure 22: Defense Spending without outliers

When looking at the temporal behavior of the scenarios generated under extreme conditions, defense spendings are again clustered as a result of combining low and high parameters values. In order to avoid cluttering of the figures, here as well the extreme cases of defense spendings with more than 46% of GDP are not shown. In each cluster, most defense spendings oscillate. Mostly, S-shaped growths can be observed. In some cases, the defense spendings decrease after a relatively steep increase.

D PRIM

When applying the PRIM algorithm to find a suitable box for the outcomes of interest as defined above, maximizing density results in the peeling trajectory that can be seen in Figure 23. The curve is concave, which means that good trade-off between density and coverage of the outcomes of interest is hard to find. The peeling trajectory begins in the lower right corner with a coverage of 1, meaning that 100% of the cases are being considered in the same box. In this very first box, the density is of 30% which means that of all the cases in this box, only a fifth is of interest. At the same time, no dimension in the uncertainty space is restricted, which means that no uncertainty can be used to describe that particular box.

The coverage decreases when moving up the peeling trajectory until a density of 85% is found. Thus, in that box, three quarters of the cases are actually of interest. However, the coverage is less than 20% and therefore, only a small number of all the cases that are of interest are actually in that box.

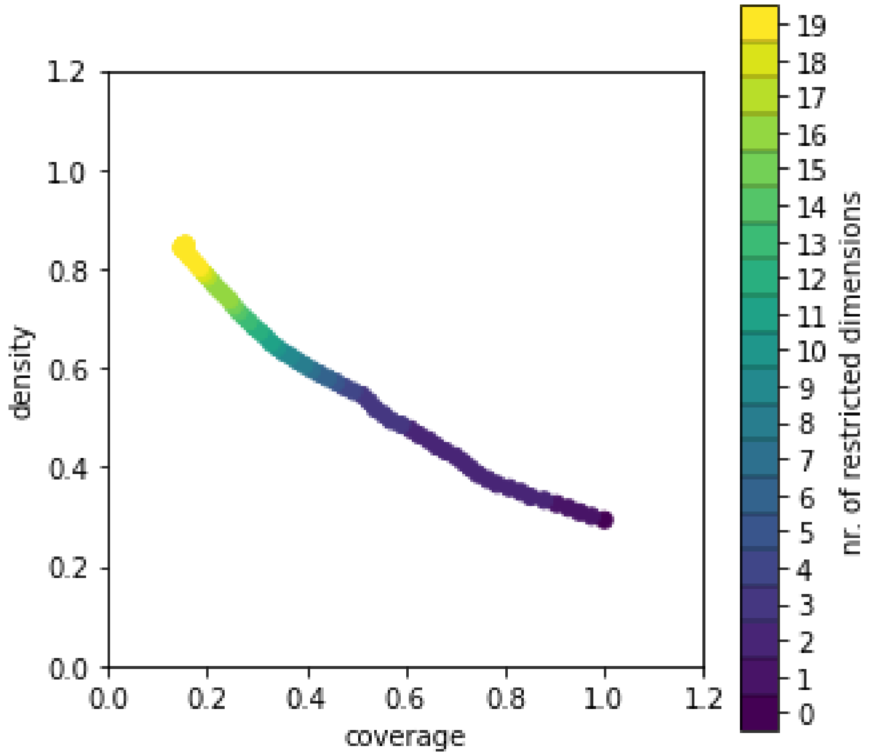
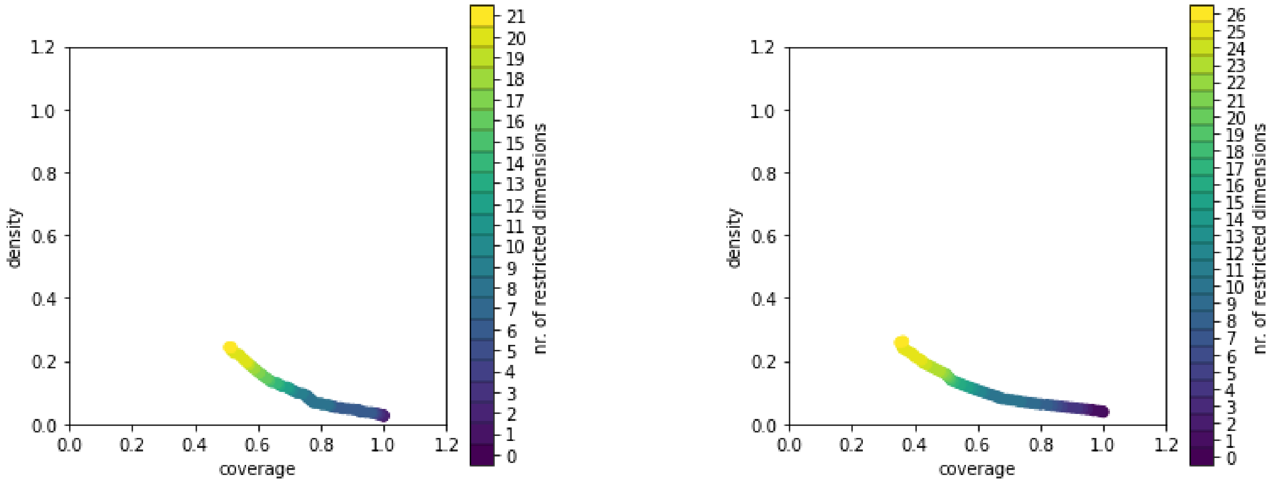


Figure 23: Tradeoff between density and coverage

Running PRIM on only those cases where the arms races are bilateral, meaning due to both nations increasing its defense spendings by 50% during a 5 year period, the results are more or less then same. The tradeoff curve between density and coverage looks relatively similar. Also, the box with a similar coverage and density as above is characterized by short response times and by low initial general capabilities and higher bias in estimating opposing autonomous capabilities for B. Running PRIM on unilateral arms races does not give any good results, as can be seen in Figure 24. As the density never exceeds 30%, the results from PRIM are as good as taking a guess.

Running PRIM on the arms races that can be found when applying the different regulations, also leads to concave curves for the density-coverage trade-off. This means that a good trade-off cannot be found, where the density and the coverage are desirable at the same time. When



(a) PRIM trade off curve for A's defense spending

(b) PRIM trade off curve for B's defense spending

Figure 24: Trade off curves density versus coverage when performing PRIM on unilateral arms races

using PRIM to identify the vulnerabilities of the best performing strategy, there is again the same difficulty.

E Clustering

Defense spendings are once clustered based on the dynamic behavior of the defense spending of nation A (see Figure 25) and once based on the one of nation B (see Figure 26).

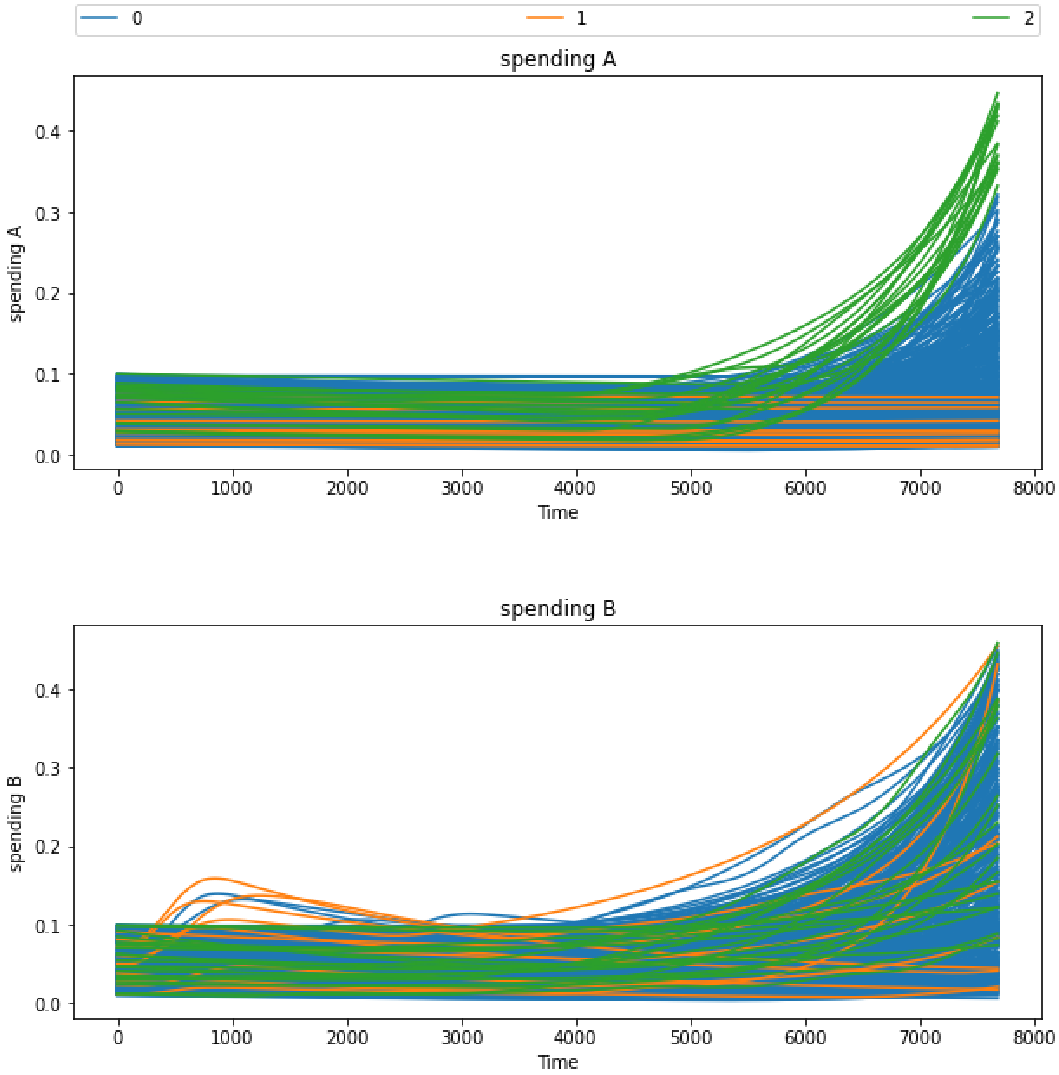


Figure 25: Clustering defense spendings based on nation A.

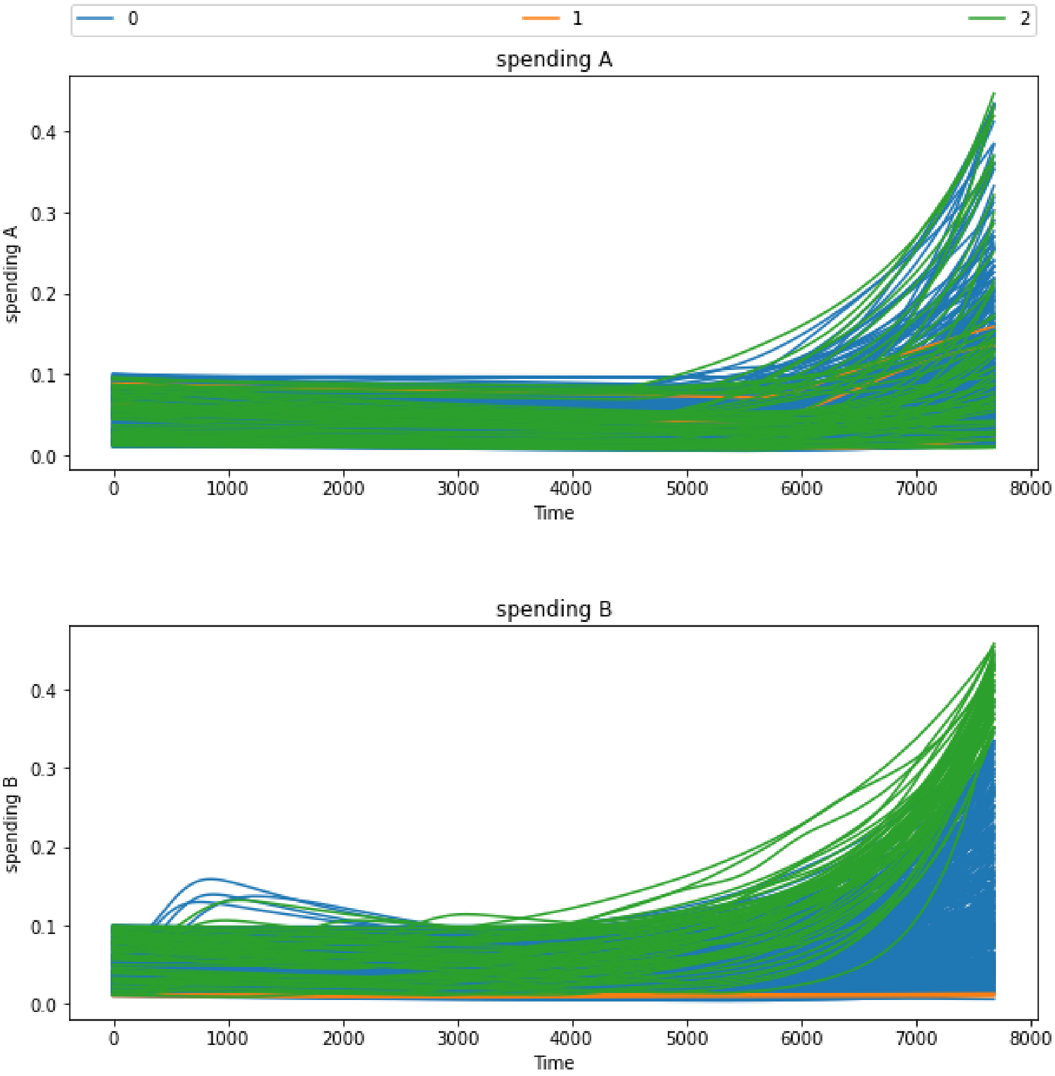


Figure 26: Clustering defense spendings based on nation B.

To begin with, when clustering based on the defense spending of hegemon nation A, the defense spending of nation A mostly behaves in an S-shaped growth (see the blue Cluster A0, making 93% of all arms races). Some of these are unilateral races that are driven only by rising nation B. In these situations, even though the defense spending of nation A behaves in an S-shaped manner, it does not increase enough for it to be characterized as an arms race. In these cases, the defense spending of B oscillates relatively often, as if nation B could not decide whether it really wants to engage in an arms race or not. In general, nation B's military power is less than nation A's, due to lower initial defense spendings and capabilities. Nation A on the other hand has a greater than average initial investment into autonomous capabilities. Hence, B is trying to keep up an engaging into an arms race. Since, due to a faster development of technological sophistication of nation B's autonomous capabilities compared to A, A will (in the bilateral cases) get involved into this arms race and increase its defense spending accordingly.

In some races, the spendings of nation A rise sharply in the last 10 years of the simulation time (green Cluster A2, being 0.05% of all arms races). This cluster also includes some unilateral arms races that are driven by nation A. In these, similarly as above, nation B's increase in its defense spending is not enough to be characterized as an arms race. These scenarios can be identified in the figure as those where the defense spendings of A increase and after reaching a global maximum, decrease again slightly towards the end of the simulation time. The main difference to the other clusters is that the initial investment into autonomous capabilities for nation B are comparably low while for nation A is higher. Mostly, B will be kicking off an arms race as it lacks behind A in military strength. Nation A then joins in as soon as nation B becomes a threat. In some cases, however, as in those where the quality of B's systems is greater than that of nation A, nation B will not get involved in an arms race as it has the technological advantage.

In some rare cases (less than 0.03%), the defense spendings of nation A remains entirely constant (see orange Cluster A1). This cluster represents the unilateral arms that are all driven by nation B. Here, nation A is investing comparably little into autonomous systems, namely 6 times lower than on average in the other scenarios. At the same time, A generally has a relatively low defense budget. Hence, with nation B having little initial autonomous capabilities, it makes sense that B will want to catch up with A. Still, this conflicts with the regular unilateral B races, where initially A invests more and B less into autonomous systems. The difference to these it that in the regular unilateral B races, nation B has initially more autonomous capabilities. As in these 1% of the cases B has initially less systems, nation A will not perceive B as a threat as long as it does not get close to A, which did not happen within the simulation time.

Moreover, when clustering based on the defense spending of nation B, there are some race cases (less than 0.01%) with constant defense expenditures of nation B (orange B1). Again, here the arms races are driven by nation A. Similarly, to above, the scenarios where B's defense spendings stay constant are characterized by an initially low defense budget of nation B. At the same time B has an advantage in developing autonomy and A invests more than average on its autonomous capabilities, while nation B invests really little. This is opposed to what is happening in the usual unilateral A driven arms races, where A initially invests less and B more into autonomous capabilities. The difference is that in these 2% special cases, nation B is very fast in developing its autonomy and hence does not perceive nation A as a threat as long as nation B keeps the edge, which it presumably does within the simulation time and hence does not have the need to increase its defense spending.

The majority of defense spending of B behave in an S-shaped growth (85% in the blue cluster B0). In this cluster, some of the arms races are actually driven only by nation A (the ones where A hits a global maximum before decreasing by the end of the simulation) since B is not increasing its expenditure enough over a long enough period of time. In these scenarios, B has an advantage in developing AI and thus autonomy. At the same time, it has low initial general capabilities and a small initial defense budget. Since A is investing a lot into autonomous capabilities, it will end up increasing its autonomous capabilities corresponding to the perceived increase in effectiveness of the autonomous capabilities of B.

Finally, 14% of B's defense spendings increase more or less sharply by the end of the simulation time (cluster B2). It also includes the scenarios in which the defense spending of B oscillates. As mentioned above, these are actually cases in which the arms race is unilaterally driven by nation B. This cluster corresponds more or less to the cluster A0 in terms of scenario characterization, with the difference of a high initial stock of autonomous capabilities for A instead of the low initial defense budget of B. In terms of mechanisms that drive the dynamics of the defense spendings, it boils down to the same: Nation B is weaker in terms of military power than nation A. Due to nation A investing a lot into autonomous capabilities, nation B needs to keep up by increasing its defense spending. The advantage of nation A is in some cases as so big that there is no need of getting involved in an arms race to stay ahead.