Panoptic Dual-Use Management: Preventing Deliberate Pandemics in an Age of Synthetic Biology and Artificial Intelligence

Jonas Sandbrink, Trinity College

August 2023

Thesis for the degree of Doctor of Philosophy

Word Count: 71,930

ABSTRACT

Panoptic Dual-Use Management: Preventing Deliberate Pandemics in an Age of Synthetic Biology and Artificial Intelligence

Jonas Sandbrink, Trinity College

August 2023

Thesis for the degree of Doctor of Philosophy

Powerful new technologies can have profound global security implications. In this thesis, I investigate how advances in synthetic biology and artificial intelligence could have dual-use potential and enable the deliberate release of pandemic pathogens. I review risks from synthetic biology based on case studies on wildlife virus discovery, viral engineering for vaccine design, and viral engineering for gene therapy. For assessing impacts of artificial intelligence, I consider large language models and biodesign tools. I find that related advances can create new methods to engineer pathogens and make such capabilities increasingly accessible to non-specialists.

These risks are not well captured by existing risk mitigation measures. I argue that the management of dual-use virological research is currently defined by oversight of individual research projects. This is effective for addressing high-risk research but fails to address risks from a more diffuse set of research and technologies with dual-use potential.

To help mitigate these risks, I introduce the idea of panoptic dual-use management. Inspired by methodologies to reduce carbon emissions, panoptic dual-use management involves treating associated dual-use risks as negative externalities and creating appropriate incentives so they are accounted for in decisions between projects. I explore ways in which such incentives could be created for various stakeholders. For instance, funding bodies could use dual-use risks as a tiebreaker between projects on the brink of getting funded, a practice which would incentivise researchers to preferentially propose projects with lower dual-use risks. To realise this proposal, I sketch out a framework for assigning tiered dual-use scores to virological research.

I conclude by highlighting the importance of combining different dual-use management approaches across stakeholders and geographies to establish an effective complex of overlapping mitigation regimes.

TABLE OF CONTENTS

INDEX OF FIGURES, BOXES, AND TABLES

Index of Figures

Figures were created with Biorender.com.

Index of Boxes

Index of Tables

ABBREVIATIONS

AAV: Adeno-associated virus

AI: Artificial intelligence

API: Application Programming Interface

BBSRC: Biotechnology and Biological Sciences Research Council (United Kingdom)

bp: base pair

BWC: Biological Weapons Convention

CBM: Confidence-building measure

CBRN: chemical, biological, radiological and nuclear

CDC: Centers for Disease Control and Prevention

CDP: Carbon Disclosure Project

DFG: German Research Foundation

DIY bio: Do-It-Yourself biology

DNA: Deoxyribonucleic acid

dsDNA: double-stranded DNA

DURC: Dual-use Research of Concern

EASAC: European Academies Science Advisory Council

ESG: Environmental, social, and governance

ePPP: enhanced potential pandemic pathogen

FDA: Food and Drug Administration (United States)

FSAP: Federal Select Agent Program (United States)

GOF: Gain-of-Function

HHS: Department of Health and Human Services (United States)

iGEM: International Genetically Engineered Machines competition

IGSC: International Gene Synthesis Consortium

kb: kilo base (1000 bases)

LLM: Large language model

MERS: Middle East Respiratory Syndrome

MRC: Medical Research Council (United Kingdom)

NIH: National Institutes of Health (USA)

P3CO: Potential Pandemic Pathogen Care and Oversight (HHS policy in United States)

PCR: Polymerase Chain Reaction

PPP: potential pandemic pathogen

RNA: Ribonucleic acid

SARS: Severe Acute Respiratory Syndrome

STS: Science and Technology Studies

UN: United Nations

US/USA: United States of America

WHO: World Health Organisation

ACKNOWLEDGEMENTS

A tremendous number of people enabled me to write this thesis. I am grateful to all of them for their support. In particular, I would like to thank my supervisors Julian Savulescu, Dominic Wilkinson, David Relman, and Kevin Esvelt, who have been very generous with their time and wisdom. Each brought their own unique perspective to my work, without which this thesis would not have been the same. This thesis would also have not been possible without the generous funding that I received from Open Philanthropy. I am indebted to Andrew Snyder-Beattie and Chris Bakerlee, who have been real champions for my work and continue to enable new thinking and progress in biosecurity.

A number of mentors, colleagues, and friends have been particularly important for this journey. My colleagues and friends Joshua Monrad and James Wagstaff have been crucial intellectual sparring partners and sources of support. The Gregs, Gregory Lewis and Gregory Koblentz, guided my very first ventures into biosecurity and continue to be generous mentors. It has been a joy to work with Shrestha Rath during the last part of this DPhil, and I am grateful for her patience in listening to my ideas and for helping out with small tasks here and there, including finding specific references for Chapters 2, 10, and 12.

I have had the pleasure of working with many remarkable individuals over the years, who shaped my thinking and writing through many discussions, emails, and written feedback. These include Janvi Ahuja, Tessa Alexanian, Ethan Alley, Markus Anderljung, Nick Bostrom, Richard Bruns, Beth Cameron, Allan Dafoe, Richard Danzig, Richard Ebright, Sam Weiss Evans, Aleš Flídr, Anemone Franz, Doug Friedman, Daniel Greene, Friederike Grosse-Holz, Hamish Hobbs, Tom Inglesby, Hannah Klim, Piers Millett, Michael Montague, Rebecca Moritz, Cassidy Nelson, Megan Palmer, Jassi Pannu, Mike Parker, Anders Sandberg, Toby Shevlane, James Smith, Damien Soghoian, Elika Somani, Claire Standley, Jacob Swett, Dev Vardeep, Andrew Weber, Nicole Wheeler, and Bridget Williams (Apologies to anyone I missed!). I owe special thanks to Nick Bostrom and my other colleagues at the Future of Humanity Institute for providing me with an intellectual harbour over the past years. And to the staff and friends at Trinity College, where I spent so many of my formative years and have found a second home.

STATEMENT OF AUTHORSHIP

This thesis draws on some previously published material. I am grateful for my co-authors on these different pieces, from whom I have learned a lot - their ideas and feedback have significantly impacted on the direction of my work. In certain places marked with a footnote (Chapters 4-6 and Chapter 11, section 11.2), I draw substantially on co-authored papers. I only draw substantially on studies where I made the predominant intellectual contribution, and I have rewritten the relevant material so that the language and all remaining errors are mine. Full details are presented in Appendix C.

Published articles which I draw on substantially in this thesis:

- **Sandbrink, Jonas B**., and Gregory D. Koblentz. 2022. "Biosecurity Risks Associated with Vaccine Platform Technologies." *Vaccine* 40 (17): 2514–23. https://doi.org/10.1016/j.vaccine.2021.02.023.
    - This paper forms the basis for Chapter 5. It has been rewritten and paraphrased to fit the context of this thesis.

- **Sandbrink, Jonas B.**, Matthew C. Watson, Andrew M. Hebbeler, and Kevin M. Esvelt. 2021. "Safety and Security Concerns Regarding Transmissible Vaccines." *Nature Ecology & Evolution* 5 (4): 405–6. https://doi.org/10.1038/s41559-021-01394-3.
    - This paper forms the basis for Chapter 5, section 5.4. It has been rewritten and paraphrased to fit the context of this thesis.

- **Sandbrink, Jonas B.**, Ethan C. Alley, Matthew C. Watson, Gregory D. Koblentz, and Kevin M. Esvelt. 2022. "Insidious Insights: Implications of Viral Vector Engineering for Pathogen Enhancement." *Gene Therapy*, March. https://doi.org/10.1038/s41434-021-00312-3.
    - This paper forms the basis for Chapter 6. It has been rewritten and paraphrased to fit the context of this thesis.

Preprints which I draw on substantially in this thesis:

- **Sandbrink, Jonas B.**, Janvi Ahuja, Jacob Swett, Gregory Koblentz, and Claire Standley. 2022. "Mitigating Biosecurity Challenges of Wildlife Virus Discovery and Characterisation." SSRN Scholarly Paper ID 4035760. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4035760.

    - This preprint forms the basis for Chapter 4. It has been rewritten and paraphrased to fit the context of this thesis.

- **Sandbrink, Jonas B.** 2023. "Artificial intelligence and biological misuse: Differentiating risks of language models and biodesign tools." arXiv preprint. https://arxiv.org/abs/2306.13952.
    - Chapter 7 forms the basis for this preprint with only minor modifications.

- **Sandbrink, Jonas B.**, Hamish Hobbs, Jacob Swett, Allan Dafoe, and Anders Sandberg. 2022. "Differential Technology Development: A Responsible Innovation Principle for Navigating Technology Risks." SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=4213670.
    - This preprint is drawn on in Chapter 11, section 11.2. It has been rewritten and paraphrased to fit the context of this thesis.

Other peer-reviewed publications and pre-prints, which I touch on:

- Musunuri, Sriharshita, **Jonas B. Sandbrink**, Joshua T. Monrad, Megan J. Palmer, and Gregory D. Koblentz. 2021. "Rapid Proliferation of Pandemic Research: Implications for Dual-Use Risks." *MBio* 12 (5): e01864-21. https://doi.org/10.1128/mBio.01864-21.
- Smith, James Andrew, and **Jonas B. Sandbrink**. 2022. "Biosecurity in an Age of Open Science." *PLOS Biology* 20 (4): e3001600. https://doi.org/10.1371/journal.pbio.3001600.
- Pannu, Jaspreet, **Jonas B. Sandbrink**, Matthew Watson, Megan J. Palmer, and David A. Relman. 2021. "Protocols and Risks: When Less Is More." *Nature Protocols*, December, 1–2. https://doi.org/10.1038/s41596-021-00655-6.

PART I: BACKGROUND

Introduction

The COVID-19 pandemic has demonstrated our society's vulnerability to biological events. In three years, COVID-19 killed over 6.5 million individuals and led to unprecedented economic costs (Our World in Data 2022). World GDP declined by 10% just in the first year of the pandemic (Harari, Keep, and Brien 2022). COVID-19 has caused immeasurable harm to mental health, education, and prospects for economic growth and raising global quality of life. Uncoordinated responses, reversion to nationalism, and abundant disinformation have demonstrated that our society is not prepared for large-scale catastrophic events of this nature.

COVID-19 has been discussed as a once-in-a-century event (Cruickshank and Shaban 2020). However, risks for pandemics might be rising. Humanity continues to spread across the globe and disrupts animal habitats, which increases the risk for the cross-species spillover of new potentially pandemic pathogens. Another important factor is the increasing risk of deliberate biological events; advances in biotechnology mean that a growing number of individuals and groups will be able to create and engineer pandemic-capable pathogens. For example, Kevin Esvelt has estimated that ten thousands of individuals are able to follow basic virus synthesis protocols (Esvelt 2022). Whether this number is accurate is contentious, but it is undeniable that the creation of viruses is becoming more accessible. If capabilities to create pandemic-capable viruses spread unfettered, at some point, risks for deliberate or accidental release might outweigh that of

natural emergence. And deliberate biological events might be much worse than naturally emerging ones: pathogens may be engineered for maximal harm or released at multiple locations.

In this thesis, I explore how new tools and methodologies to tackle infectious diseases or develop new therapeutics could inform the creation and deliberate release of pandemic-capable pathogens - and what society could do to manage this dual-use potential of synthetic virology.

My interest in this topic arose over the course of my studies and previous research. I studied medicine during a global pandemic. Caring for COVID-19 patients on Oxford's intensive care unit made me viscerally appreciate the grave costs of infectious disease. Engaging in epidemiological research on interventions like lockdowns made me realise society's inability to respond effectively to health crises. Even before COVID-19, I was interested in pandemic prevention and had done research on RNA vaccines. My concern for biological events and knowledge of cutting-edge biotechnology inspired me to consider the dual-use risks of advances in viral engineering. My long-standing interest in the impacts of artificial intelligence and the publication of tools like chatGPT made this an obvious area to include as well. I soon realised that existing oversight failed to consider a substantial fraction of what I perceived as research with noteworthy dual-use risks.

One example of a paper that made me pause was a study to improve adeno-associated virus (AAV) for gene therapy delivery. In this study, Chan *et al.* describe their experiments to insert short non-coding DNA sequences into AAV to prevent the activation of the immune response (Chan et al.

2021). Specifically, the authors tested these short DNA inserts to interfere with the activation of the TLR9 innate immune sensor, which normally detects viral genetic material and subsequently induces an inflammatory response. These small non-coding DNA fragments constitute universal genetic elements that could be inserted into a wide range of pathogens. For this specific study, the immune modulation effect is still limited. However, future iterations to improve immune modulation could result in a straightforward method to make almost any pathogen evade immune activation.

I reflected on studies similar to Chan *et al.*. I could see the benefits of exploring methods for advancing gene therapy, and appreciated that risks were remote and indirect. However, I also realised that many alternative approaches without the same risks were available to achieve the same goal. Such alternatives include vector-specific immune evasion methods or the use of immune-modulatory medications during the administration of gene therapy. I had looked into existing efforts to address dual-use risks and knew that existing oversight was narrowly focussed on experiments with specific pathogens - and thus, the line of inquiry for studies like Chan *et al.* was likely never subjected to risk evaluation. I did not know how oversight structures could be adjusted to address these diffuse dual-use risks - but it got me thinking. Only later, I considered parallels in climate change and realised how some of the lessons from decarbonisation might apply to dual-use risk mitigation. This thesis draws together my thinking on the dual-use risks of studies such as Chan *et al.*, how these studies are changing the landscape of dual-use research, and how existing oversight approaches could evolve to meet the challenge.

In Part I (Chapters 1-3), I lay out the background for my subsequent analysis. In Chapter 1, I discuss the history of biological weapons and high-profile debates on dual-use research, a history which, among other things, highlights that actors interested in the catastrophic misuse of biotechnology exist. Then, in Chapter 2, I provide an overview of how dual-use risks are considered in different countries and research communities, including among life scientists, sociologists, and ethicists. In the last chapter of the background section, Chapter 3, I dive into the scientific advances that are shaping emerging dual-use risks.

In Part II (Chapters 4-7), I analyse emerging dual-use risks in synthetic virology and artificial intelligence. I evaluate a series of research areas for emerging dual-use risks: discovery of viruses in wildlife (Chapter 4), viral engineering for vaccines (Chapter 5), and viral engineering for gene therapy (Chapter 6). I also evaluate the impacts of artificial intelligence, such as large language models and biodesign tools (Chapter 7). For each of these areas, I analyse how advances might make the creation of pandemic-capable viruses more accessible and how scientific approaches could be used to mitigate risks. I consider how researchers might develop technologies in a way that prevents misuse or whether lower-risk alternatives exist that would offer the same level of benefits.

In Part III (Chapters 8-10), I put emerging dual-use risks in the context of existing dual-use oversight approaches. In Chapter 8, I discuss how emerging dual-use risks challenge existing oversight policies of the United States, policies which have been influential globally. In Chapter 9, I zoom in on a concrete lesson that also the Chan *et al.* study highlights: increasingly,

transferable insights and general-purpose methods for viral engineering are driving dual-use viral engineering capabilities. A broad set of research with indirect dual-use risks will be sufficient to enable a growing number of individuals the ability to start a pandemic. Thus, I argue that dual-use management needs to consider this broader set of research that advances transferable insights and general-purpose approaches. However, existing dual-use risk mitigation approaches are not designed to do so. In Chapter 10, I characterise the current approach of overseeing individual projects and analyse how it could be strengthened. However, I find that an approach focused on individual projects ultimately has limits for addressing increasingly diffuse dual-use risks.

In Part IV (Chapters 11-13), I propose dual-use management approaches suited for the changing risk landscape, a risk landscape where misuse is increasingly enabled by a more diffuse set of transferable insights and general-purpose tools. In Chapter 11, I introduce the concept of panoptic dual-use management. Inspired by the Greek *panoptes*, meaning 'all-seeing', panoptic dual-use management involves consideration of all possible research projects and factoring dual-use risks as negative externalities into decisions between them. This may require creating appropriate incentives for stakeholders. As I develop the concept of panoptic dual-use management, I consider lessons from climate change mitigation on how to address negative side-effects associated with a large set of beneficial technologies. Decarbonisation required looking towards low-emission alternatives and aligning incentives to account for the societal costs of carbon emissions. I argue that these approaches are transferable to mitigating the dual-use risks of synthetic virology. In Chapter 12, I then analyse how policymakers and other stakeholders could implement panoptic dual-use management. I put forward ideas for how different stakeholders

could be incentivised to consider dual-use risks in decision-making on research projects. I flesh out one concrete such idea in Chapter 13: funding bodies could use dual-use risks as a tie-breaker between projects on the brink of being funded. I highlight how this approach could be implemented with minimal overhead and, to this end, propose a framework to assign tiered dual-use scores for synthetic virology research.

In Part V (Chapter 14), I draw all my findings together. I highlight that just like no single regime is used to tackle decarbonisation, no single approach will solve the dual-use problem. Rather, different approaches implemented at different levels could come together to form a regime complex that, overall, is steering away from a future with widely accessible capabilities for mass destruction. I also consider overarching challenges to advancing dual-use risk mitigation, including the issue of global coordination. Lastly, I explore the larger strategic picture for preventing catastrophic misuse. I argue that dual-use risk management won't be able to prevent the misuse of certain capabilities forever; thus, robust capabilities to contain a pandemic will be crucial for closing the risk window for the catastrophic misuse of biotechnology.

This work touches on a great number of important disciplines and topics - and I cannot hope to do justice to all of them. A non-exhaustive list of areas I discuss includes virology, public health, bioethics, artificial intelligence, Science and Technology Studies, public policy, economics, and sustainability. Where I draw on a discipline, I tried to research crucial background information and references to apply its perspective to the best of my knowledge. However, I expect I have not done complete justice to all of them. I ask for the reader's forgiveness if they find any crucial perspectives from their discipline insufficient or missing.

Addressing dual use risks associated with advances in virology will require input, expertise and insights from multiple different disciplines. This thesis is in no way a final word on this problem; rather, I see it as another step towards a broader conversation about how society can mitigate risks that accompany scientific and technological advances.

Methodological note on information hazards

With this thesis, I hope to shed light on emerging risks that help inform risk mitigation. However, any work to characterise threats could also inspire and inform misuse. Thus, I am mindful of information hazards[1] associated with my own writing. I tried to balance the risks and benefits of disclosing possible information hazards. When planning my chapters, I considered which topics and modes of analysis would favour risk mitigation over misuse. For instance, I ended up scrapping a breakdown of all the different ways in which a pandemic-capable virus could be created. Publishing such a list in a thesis that will be publicly accessible would potentially provide a recipe book for those who might wish to engineer a bioweapon. When deciding whether to disclose a concrete piece of dual-use information, I considered the magnitude of its information hazard potential and the importance of disclosure for risk mitigation.[2] Where possible, I tried to use examples with less direct misuse potential to make a given point. For instance, I carefully selected the Chan *et al.* study as a prominent example to illustrate my arguments. Indeed, a different study initially sparked my concerns and inspired much of this thesis; however, I felt its misuse potential was too imminent and that it would be a mistake to bring much attention to it. I cross-checked crucial judgements on disclosing information hazards with at least one colleague.

[1] While the idea of harmful knowledge is an ancient trope, the academic concept of an "information hazard" was first defined by Nick Bostrom. He gives the following definition: "A risk that arises from the dissemination or the potential dissemination of (true) information that may cause harm or enable some agent to cause harm." (Bostrom 2012)

[2] For the magnitude of information hazard potential I evaluated the novelty of the information, the associated scale of misuse, and how directly and imminently it could be misused. For the potential to contribute to a solution, I considered the likelihood of a solution existing, how dependent such a solution would be on disclosure, and how likely a less actionable disclosure by someone else would be. Where I disclosed possible information hazards, I tried to give the least dangerous example with minimal necessary detail, and tried to frame disclosure in a way that would maximise success for a solution. More details on the framework I applied can be found in Appendix A.

Chapter 1: The history of dual-use

Dual-use research is research conducted for legitimate purposes that has the potential to be misused. Researchers and other stakeholders interested in advancing such research thus face a "dual-use dilemma" because their work may enable misuse resulting in harm (S. Miller and Selgelid 2007). A well-known example of the dual-use dilemma is the enhancement of potential pandemic pathogens:

> A researcher hopes to study whether an avian influenza virus might be able to cause a future pandemic. Avian influenza, colloquially known as bird flu, occasionally jumps from birds into humans. For H5N1, the subtype of avian influenza virus the researcher is interested in, the World Health Organisation estimates that 60% of infected humans die (Li et al. 2008). The researcher is concerned that this virus might acquire genetic changes that make it human-to-human transmissible. She considers conducting an experiment to identify such genetic changes and thus evaluates the benefits and risks of this work. Identifying genetic changes for human-transmissible avian influenza could help inform disease surveillance and vaccine development, and, therefore, might help to prevent a pandemic. At the same time, identifying such genetic changes might provide bioterrorists-to-be with a blueprint for a pathogen capable of causing a global catastrophe.[3] How should she trade off these benefits and risks? Should she conduct the experiment or not?

In practice, scientists or funders rarely decide not to conduct a given study but rather modify experiments or adopt communication strategies to mitigate risks (S. W. Evans et al. 2021). Thus, decisions about dual-use research do not necessarily come down to a dilemma of two choices but rather a range of risk mitigation

[3] There are also safety and physical security risks of conducting an experiment to enhance a potential pandemic pathogen, which indeed have dominated many related discussions; however, I have simplified the discussion of risks and benefits to highlight the dual-use dilemma.

strategies. Therefore, the "dual-use dilemma" may be better described as a "dual-use problem" (Douglas 2013). In this thesis, I use the more general framing of research featuring "dual-use risk" or "dual-use potential", risks mitigated through strategies that modify the research or refrain from the research in question in favour of pursuing alternatives.

"Dual-use" has been used in many different contexts. "Dual-use" can refer to technologies that have both benevolent and harmful applications, military and non-military applications, or offensive and protective applications within a military context (S. Miller and Selgelid 2007). The enhancement of potential pandemic pathogens is described as "dual-use" in the sense of benevolent/harmful. In contrast, biodefense research on pathogen dispersal could be considered "dual-use" in the sense of offensive/protective. As I mainly talk about non-military research and instances of misuse by rogue individuals or groups, I use "dual-use" in the former sense of beneficial/harmful.

Dual-use risks may present across different dimensions of research. Atlas and Dando describe three aspects of the dual-use problem: 1) the misuse of civilian facilities, 2) the misuse of equipment and agents, and 3) the misapplication of knowledge and insights (Atlas and Dando 2006). Arguably, the misuse of civilian facilities mostly applies to state biological weapons programs; thus, I focus less on this aspect of the dual-use problem. I sometimes touch on laboratory biosecurity, such as the misuse or theft of agents and other physical research products. Most of my discussion focuses on managing dual-use risks of scientific knowledge and insights. As synthetic biology is becoming increasingly accessible, Lewis *et al.* assign the majority of security concerns to such intangible research products (Lewis et al. 2019). Whether intangible research products like knowledge and insights actually pose the majority of misuse risk or not, they are likely one important risk factor. Throughout this thesis, I differentiate between biosecurity and biosafety

risks. Biosecurity risks capture intentional misuse, including the misapplication of dual-use insights but also laboratory biosecurity, while biosafety risks refer to accidental pathogen exposure or release.

Existing dual-use management is shaped by historical incidents of biological weapons use and dual-use research. This chapter aims to provide an overview of these events. First, I recount historical biological weapons efforts and present known bioterrorism attempts (1.2). Then, I summarise debates on specific examples of dual-use research and the policies they have shaped, examples and policies that I refer to throughout my thesis. I first discuss debates of the 2000s (1.3) and follow with more recent discussions, including debates about the enhancement of potential pandemic pathogens (1.4).

1.1 Biological weapons

1.1.1 State biowarfare

Discussions on the dual-use potential of life sciences research are framed by the development and use of biological agents throughout history. The history of biological agents goes back to before the scientific understanding of disease (Carus 2017). The first well-documented offensive use of a biological agent occurred in 1763 when the British handed Native Americans blankets from smallpox hospitals to start an outbreak and create a military advantage (Ranlet 2000).

Modern, science-informed biological warfare started around World War I. Germany was the first country to have a biological warfare program. The program focused on animal pathogens, including *Bacillus anthracis*, the agent that causes anthrax. With these pathogens, Germany sabotaged supply lines and munitions factories, including in the United States, which was neutral at the time (Wheelis 1998). France may have also engaged in sabotage using biological weapons during World War I. After the war, in 1925,

the Geneva Protocol banned the use of chemical and biological weapons; however, this treaty did not prohibit possession nor use for retaliation (Thomas 1970). Over the decades following the war, many countries started biological warfare programs of varying scope, including Hungary, Italy, Japan, Poland, and the Soviet Union (Carus 2017).

World War II brought the largest-ever use of biological weapons. Japan's biological weapon program was the most notorious program of World War II and is estimated to have killed tens of thousands of Chinese during the war (Carus 2017). Most casualties were inflicted by fleas infected with *Yersinia pestis*, the agent causing the plague. The Japanese program is most infamous for its human experimentation on Chinese prisoners (Keiichi and Junkerman 2013; Carus 2017).

During the Cold War, the United States and the Soviet Union invested heavily in biological warfare capabilities. These programs turned biological warfare agents for the first time into weapons of mass destruction (Carus 2017). However, neither power ever deployed these capabilities for large-scale warfare. The United States had a technologically advanced program capable of delivering lethal biological agents using spray tank systems on bombers (Carus 2017). United States President Richard Nixon terminated the program in 1969 after judging that it did not add substantially to United States military capabilities (Tucker and Mahan 2009).

During the Cold War, the Soviet Union organised the largest-ever biological weapons program, with an estimated 60,000 employees at its peak (Carus 2017). Among many other activities, researchers in the Soviet bioweapons program engineered viruses in a highly sophisticated manner, creating chimeric viruses incorporating genes from different organisms and inserting bioregulatory genes to disrupt the immune system (Gilsdorf and Zilinskas 2005). The program started in the 1920s and stayed active throughout the

Soviet Union's dissolution, despite the Soviet Union leading international negotiations on the Biological Weapons Convention in the 1960s and 1970s (Leitenberg and Zilinskas 2012).

The Biological Weapons Convention (BWC) was a significant biological arms control success. Signed in 1972 and entering into force in 1975, it was the first treaty to ban a whole class of weapons (UNODA 2022). Despite lacking a mechanism to verify and enforce compliance, the treaty has created strong norms against the military use of biological agents. However, the Soviet Union is not the only known violator of this treaty. South Africa's apartheid regime developed biological weapons in the 1980s, mainly focussing on assassination agents and exploring antifertility drugs for South Africa's black population (Gould et al. 2002). In 1983, Saddam Hussein revived Iraq's biological weapons program (Carus 2017). The program involved filling missiles with various biological agents. However, Hussein never used these weapons in the Gulf War of 1991 due to uncertainty over their effectiveness. After 1991, United Nations inspections investigated Iraq's biological weapons, finding evidence of work on bacteria and toxins as well as camelpox virus, Enterovirus, and rotavirus (Central Intelligence Agency 2005). It is unclear which countries retain biological warfare programs. The United States judges that Russia and North Korea continue to have an offensive biological weapons program and has concerns that China and Iran are conducting activities which violate the Biological Weapons Convention (Department of State 2016; Harris 2020).

1.1.2 Bioterrorism

While state programs define the history of weaponised biology, dual-use management is predominantly shaped by historical bioterrorism attempts. Non-state actors that have tried to weaponise biology range from small, well-resourced groups to opportunistic individuals. One of the most notorious terrorist groups exploring biological weapons was the Japanese doomsday cult Aum Shinrikyo. Starting in 1990, members of Aum Shinrikyo attempted to produce *B. anthracis* and *botulinum toxin*. Seiichi Endo, a Kyoto

University-trained virologist, led Aum Shinrikyo's attempts to weaponise biology (Danzig et al. 2012). In this work, Endo could draw on Aum Shinrikyo's significant budget of 10s of millions of US dollars (Leitenberg 1999). Over four years of research, Endo and his colleagues successfully obtained a vaccine strain of *B. anthracis,* produced significant amounts of this agent, and disseminated it. The attacks were unsuccessful as they had failed to render the vaccine strain pathogenic.

Another terrorist group that attempted to weaponise biology is Al-Qaeda. Al-Qaeda started these attempts in the 1990s but failed to make significant progress by the September 11 attacks of 2001. The United States disrupted these efforts when invading Afghanistan (Leitenberg 2005). Notably, Al-Qaeda was motivated in part to start the program by the US government publicly pointing towards the threat of weaponised biology (Wright 2002). Thus, discussion of misuse risks can have dual-use risks in itself if it inspires or informs malicious actors.

Public awareness of dual-use risks was heightened by a bioterror attack perpetrated by a single individual. In the week following 11 September 2001, refined *B. anthracis* was sent by mail to multiple public figures in the United States. Five people were killed, 17 injured. These anthrax attacks, known as Amerithrax, were investigated for many years by the FBI. Eventually, they were traced back to Bruce Ivins, an anthrax researcher at the US Army Research Institute of Infectious Disease (Guillemin 2011).

Based on these historical cases of bioterrorism, efforts to govern the dual-use potential of life sciences research mainly aim to prevent misuse by small ideological groups and single malicious individuals. There are multiple reasons to focus on sub-state actors. First, non-state actors may be ideologically motivated to cause the greatest potential harm and thus may be particularly likely to weaponise pandemic viruses. Second, non-state actors are generally more restrained by technical capabilities than states looking to

weaponise biology - and thus, dual-use management may actually impede them from achieving their goals. In my analysis, I focus on preventing misuse by non-state actors, with a particular eye on more sophisticated actors able to draw on formal molecular biology training and significant financial resources.

1.2 Debates of the 2000s and the emergence of DURC

The Amerithrax attacks sparked fear of bioterrorism in the United States, stoking the debate around the dual-use potential of life sciences research and large-scale government spending on biodefense (Franco and Sell 2011). Partly because of the drive towards national security post 9/11, the United States has been the epicentre of discussions on the misuse risks of cutting-edge life sciences. United States dual-use debates and policies shape global attitudes and governance approaches. For this reason, I will focus particularly on developments in the United States. An overview of important events is shown in Figure 1.1.

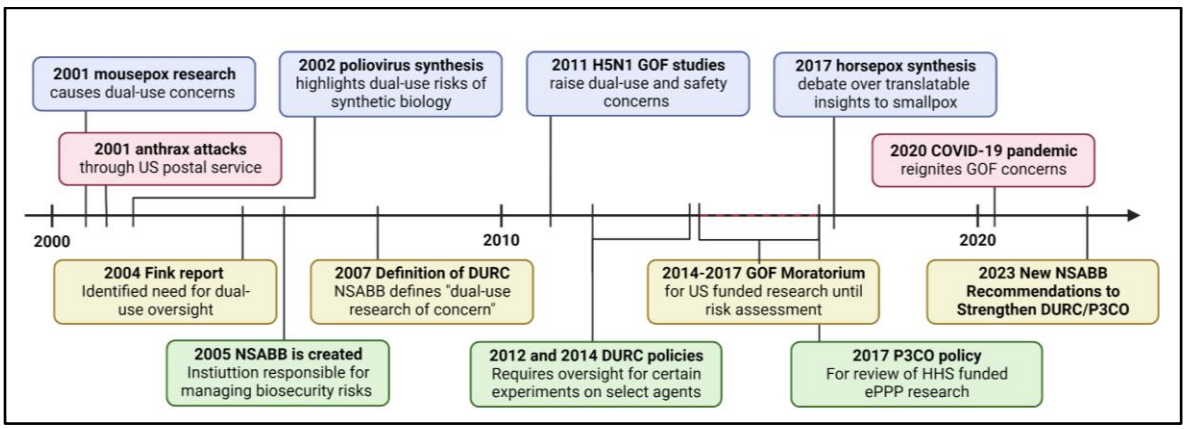


Figure 1.1: A timeline of United States dual-use debates and responses

This timeline shows events with significance for dual-use management from 2000 to 2023. In blue are examples of research projects that sparked debate over dual-use risks. In red are biological events that sparked debate over research risks. In yellow are intermediate responses by the United States government to address concerns and advance discussions. In green are decisions by the United States government for long-term dual-use management.

GOF = gain-of-function research; DURC = dual-use research of concern; HHS = US Department of Health and Human Services; ePPP = enhancement of potential pandemic pathogens; NSABB = National Science Advisory Board on Biosecurity.

On top of the Amerithrax attacks, in the early 2000s, a series of papers with potential for misuse kicked off an in-depth consideration of the dual-use risks in the life sciences. In 2001, Australian researchers published experiments that showed that inserting a genetic element encoding the immune-modulating inflammatory regulator IL-4 into mousepox virus conveyed immune evasion (R. J. Jackson et al. 2001). The researchers intended to advance rodent pest control measures. However, critics argued that this research might inform the enhancement of related variola virus, the agent that causes smallpox (Finkel 2001; Selgelid and Weir 2010). In 2002, only one year later, another paper sparked fears about the enhancement of variola virus (Rosengard et al. 2002). The same year, the virologist Eckard Wimmer published the first successful synthesis of poliovirus from synthetic DNA (Cello, Paul, and Wimmer 2002). This work heralded a new age of synthetic virology, an era in which researchers can synthesise viruses from DNA (see Chapter 3).

In response to these dual-use studies, the United States National Academies of Sciences engaged in early discussions on addressing the misuse potential of life sciences research, culminating in the publication of the influential Fink report in 2004. This report, titled "Biotechnology Research in an Age of Terrorism", identified a need for dual-use oversight, defined seven categories of experiments of concern deserving special attention, and sparked the 2005 creation of the National Science Advisory Board on Biosecurity (NSABB) (National Research Council 2004).

Over the subsequent years, political factors challenged the ambitious goals and broad remit of the Fink report. For instance, while the Fink report envisioned the NSABB as an independent entity, the NSABB

ended up being integrated into the National Institutes of Health, one of the main targets of its oversight. At the same time, NSABB being a government agency means it cannot effectively advise outside entities, like academic journals (Casadevall et al. 2014). In 2005, the publication of the genetic blueprint for the influenza virus of the 1918 pandemic in *Science* sparked security concerns (Taubenberger et al. 2005; Tumpey et al. 2005). At the last minute, the publication was reviewed by the NSABB and approved. However, in a commentary *Science* Editor-in-Chief said they would have published the paper regardless of what NSABB said unless the information had been classified (Kennedy 2005).

In 2007, the NSABB identified research sufficiently concerning to require oversight as Dual-use Research of Concern (DURC). NSABB defined DURC as:

> *"Research that, based on current understanding, can be reasonably anticipated to provide knowledge, products, or technologies that could be directly misapplied by others to pose a threat to public health and safety, agricultural crops and other plants, animals, the environment, or materiel." (National Science Advisory Board for Biosecurity 2007).*

The NSABB identified DURC based on the seven categories of experiments of concern identified by the Fink report. While the Fink report had highlighted that other research could also pose significant risks, the NSABB proposed to limit dual-use review to DURC.

Eventually, in 2012 and 2014, two Department of Human Health and Services policies for DURC oversight were established. The 2012 policy required federal funding institutions to review DURC, while the 2014 policy required all federally funded institutions to assess for DURC (U.S. Department of Health and Human Services 2012; 2014). These policies defined DURC even more narrowly than the 2007 NSABB paper: DURC review only applied to research involving the seven classes of experiments if performed on fifteen select agents and toxins (see Tables 1.1 and 1.2,). Additionally, the policies

encouraged individual investigators to raise concerns about research that falls outside these categories on a voluntary basis.

Table 1.1: Select Agents and Toxins captured by US DURC policies

Select Agents and Toxins	
Avian influenza virus (highly pathogenic)	Marburg virus
Bacillus anthracis	Reconstructed 1918 Influenza virus
Botulinum neurotoxin	Rinderpest virus
Burkholderia mallei	Toxin-producing strains of Clostridium botulinum
Burkholderia pseudomallei	
Ebola virus	Variola major virus
Foot-and-mouth disease virus	Variola minor virus
Francisella tularensis	Yersinia pestis

Table 1.2: Classes of Experiments captured by US DURC policies

Classes of Experiments
Enhances the harmful consequences of the agent or toxin
Disrupts immunity or the effectiveness of an immunisation against the agent or toxin without clinical and/or agricultural justification

Confers to the agent or toxin resistance to clinically and/or agriculturally useful prophylactic or therapeutic interventions against that agent or toxin or facilitates their ability to evade detection methodologies
Increases the stability, transmissibility, or the ability to disseminate the agent or toxin
Alters the host range or tropism of the agent or toxin
Enhances the susceptibility of a host population to the agent or toxin
Generates or reconstitutes an eradicated or extinct agent or toxin listed in Table 1.1

The United States Government instituted the DURC policies not only to mitigate risks but also to "collect information needed to inform the development of an updated policy, as needed, for the oversight of DURC". Nevertheless, these policies have not been updated to date. However, in March 2023, the NSABB released recommendations about how to revise existing US biosecurity policies (National Science Advisory Board for Biosecurity 2023). These recommendations include substantial changes, such as expanding DURC beyond the list of select agents (Table 1.1), extending it beyond federally funded institutions to all relevant research in the United States, and consolidating DURC with more recent P3CO policies on oversight of enhanced potential pandemic pathogens (discussed in the next section). I will discuss the new NSABB recommendations and how they address weaknesses of the current DURC policies in Chapter 8.

1.3 Debates of the 2010s and GOF research

1.3.1 The gain-of-function debate

In 2011, just before the publication of the first DURC policy, a new debate on the risks of life sciences research shook the global scientific community. In 2011, two groups led by Ron Fouchier (Netherlands) and Yoshihiro Kawaoka (US) presented experiments involving the enhancement of highly pathogenic H5N1 avian influenza (Herfst et al. 2012; Imai et al. 2012). This virus occasionally infects humans, killing more than half of the known infected. To identify what would be needed to make such a virus transmissible in humans, Fouchier serially passed the avian influenza virus between ferrets, a popular animal model for human influenza infections. Thus, Fouchier selected for mutations rendering H5N1 transmissible in mammals, genetic changes which he details in the publication of the work (Herfst et al. 2012). So-called gain-of-function (GOF) research, the enhancement of potential pandemic pathogens, had already in 2004 led to initial debates about how to trade-off its benefits and risks (Enserink 2004). The scenario at the beginning of this chapter, in which a researcher is considering characterising genetic changes of avian influenza, captures the difficult questions this research raises.

Initially, the United States NSABB recommended against the full publication of the Fouchier and Kawaoka studies due to the misuse potential of this work. However, after a high-profile World Health Organisation (WHO) consultation of experts, NSABB overturned its initial decision, leading to the 2012 publication of the studies (World Health Organisation 2012; Selgelid 2013).

Nevertheless, the publication of these studies had a global impact. Virologists voluntarily paused similar GOF research for 60 days (Fouchier et al. 2012). After the WHO consultation, this voluntary moratorium was extended to one year (World Health Organisation 2012; Patterson et al. 2013). This year was supposed to allow the WHO, governments, and funding bodies to review risks, evaluate additional biosafety

precautions, and create appropriate funding policies (Patterson et al. 2013). Ensuing discussions were not just limited to the risks and ethics of GOF research, but dual-use research more broadly, including at the WHO and in the European Union (World Health Organisation 2013; Fears and ter Meulen 2015).

In 2014, new developments spurred the debate in the United States. Biosafety protocol breaches at the Centre for Disease Controls and Prevention (CDC) and the Food and Drug Administration (FDA)/NIH demonstrated the risk of laboratory accidents. These safety breaches brought on calls for a longer moratorium (Burki 2018). 200+ scientists argued in the Cambridge Working Group declaration for a moratorium until a proper risk-benefit assessment had occurred. In October 2014, a federal moratorium on GOF experiments on influenza, MERS, and SARS viruses was put in place. This moratorium paused work on 18 projects "until a robust and broad deliberative process is completed that results in the adoption of a new US Government gain-of-function research policy" (Burki 2018). The moratorium was criticised for excluding work to characterise zoonotic spillover risk, including risk characterisation of animal coronaviruses.

The process of arriving at a new policy took until 2017, as different informative studies were contracted: In 2016, Gryphon Scientific published a government-commissioned analysis of the risks and benefits of Gain-of-Function research. The report made an attempt at quantifying risks and found this to be challenging; the quantification of benefits was found to be even more difficult (Gryphon Scientific 2016). Additionally, Michael Selgelid conducted a government-commissioned ethical analysis of GOF research (Selgelid 2016). Based on these commissioned studies, NSABB created recommendations for a new oversight framework (National Science Advisory Board for Biosecurity 2016), which informed the creation of White House Office of Science and Technology Policy (OSTP) guidance. In turn, this informed a new policy of the Human Health and Services department (HHS).

In 2017, HHS completed its review process and created the new P3CO framework for guiding funding of GOF research, which was relabelled "enhancement of potential pandemic pathogens" (U.S. Department of Health and Human Services 2017). With its publication, the moratorium on GOF research ended. The new HHS P3CO framework defined "potential pandemic pathogens" (PPPs) and "enhanced potential pandemic pathogens" (ePPP) to which it would apply (Box 1.1). The framework required that NIH would have to refer cases of ePPP research to HHS, where decisions would then be made on a case-by-case basis by a multidisciplinary review board. The P3CO policy was implemented in addition to the existing DURC policies.

Box 1.1: P3CO definitions and scope

Quoted from P3CO policy (U.S. Department of Health and Human Services 2017)

A. A *potential pandemic pathogen (PPP)* is a pathogen that satisfies both of the following:
    1. It is likely highly transmissible and likely capable of wide and uncontrollable spread in human populations; and
    2. It is likely highly virulent and likely to cause significant morbidity and/or mortality in humans.

B. An *enhanced PPP* is defined as a PPP resulting from the enhancement of the transmissibility and/or virulence of a pathogen. Enhanced PPPs do not include naturally occurring pathogens that are circulating in or have been recovered from nature, regardless of their pandemic potential.

C. To the extent that transmissibility and/or virulence of PPPs are modified in the following categories of studies, the resulting pathogens are not considered to be enhanced PPPs for the purposes of this Framework:
    1. Surveillance activities, including sampling and sequencing; and
    2. Activities associated with developing and producing vaccines, such as generation of high growth strains.

D. Proposed intramural and extramural life sciences research that is being considered for funding and that has been determined by the funding agency as reasonably anticipated to create, transfer, or use enhanced PPPs is subject to additional HHS department-level review as outlined herein.

E. A pathogen previously considered by an agency to be an enhanced PPP should no longer be so considered if the HHS [...] [and others decide] that the department-level review processes outlined in this framework are no longer appropriate.

While the P3CO framework stepped beyond the DURC policies by having a focus broader than select agents, it still has been criticised to be lacking in different ways. P3CO is limited to funders under the HHS umbrella and does not capture other publicly[4] or privately funded research (Inglesby and Lipsitch 2020).

[4] It has been suggested to me in personal communication that there may be a blanket ban on funding of relevant research for agencies that do not have a P3CO-like review process.

Furthermore, transparency of approvals and HHS decisions is lacking (Inglesby and Lipsitch 2020). Additionally, there is not much guidance on working with international researchers. Like the GOF moratorium, P3CO features an exclusion cause for research that modifies the transmissibility and virulence of PPPs for surveillance activities. The NSABB picked up all of these criticisms in its March 2023 recommendations for how to strengthen US Government biosecurity policies (National Science Advisory Board for Biosecurity 2023). Lastly, one fundamental property of the P3CO framework (as for the DURC policies) is that it focuses on evaluating risks and benefits of individual proposals. Thus, it could potentially fail to sufficiently consider the role of alternative experiments that may provide a superior expected net benefit (N. G. Evans 2018). Whether and how such an individual proposal-focused approach limits dual-use management, I will consider in Chapter 11 of this thesis.

1.3.2 Horsepox synthesis and other recent debates

In 2017, the synthesis of horsepox virus again highlighted the dual-use potential of life sciences research. The work was sponsored by the US biotechnology company Tonix and presumably conducted to inform new vaccines (Noyce, Lederman, and Evans 2018). Horsepox virus is closely related to variola virus, which causes smallpox. After the eradication of smallpox in 1980, vaccination programs were stopped; thus, little or no immunity against smallpox remains today (Strassburg 1982). This makes a possible release of variola virus very dangerous (Inglesby 2018). To prevent such a reemergence, only two official repositories worldwide hold variola virus, and all experiments need to be approved by a WHO committee (Centers for Disease Control and Prevention 2014). However, the full genome of variola virus was published in 1994 and is available online (Massung et al. 1994). With the advent of the possibility to create viruses from synthetic DNA (see 1.4.1), this genome now may serve as a blueprint for the resurrection of smallpox. The risk of malicious recreation of smallpox is currently low, as variola virus has a large genome and is thus difficult to synthesise. However, after the publication of the horsepox synthesis, experts feared that the

described methods lowered the barrier for the malicious resurrection of variola virus (Koblentz 2017; Inglesby 2018).

Many experts argued the risks outweighed the benefits of this work and highlighted lessons for dual-use management (Koblentz 2018; Esvelt 2018; Inglesby 2018). Tom Inglesby of Johns Hopkins University argued the promised benefit of using horsepox as a vaccine is questionable, and the publication of the details of its synthesis was completely unnecessary for this goal (Inglesby 2018). He highlighted that showcasing the possibility of a scientific capability, which was stated as a benefit of this work (Kupferschmidt 2017), is not an acceptable justification for dual-use research. Inglesby also demanded more transparency around this kind of research, which in this case could have led to the CDC sharing a horsepox strain which would have made the synthesis experiments unnecessary (Inglesby 2018).

Uncertainty around the origins of SARS-CoV-2 again increased public scrutiny of the enhancement of potential pandemic pathogens (Zimmer and Gorman 2021). In 2021, NIH Director Francis Collins was asked by US Senator Grassley to provide details about the funding of coronavirus research at the Wuhan Institute of Virology (WIV). In response, Collins stated that the relevant funding was exempt under the 2014-2017 GOF funding moratorium. Collins argued that the creation of chimeric coronaviruses which took place at WIV was at the time not considered an enhancement of potential pandemic pathogens and was covered by the moratorium's exclusion of surveillance activities (F. Collins 2021). This same exclusion clause is still part of P3CO (see Box 1.1), and thus the creation of potential high-risk chimaeras from viruses discovered in animals continues without risk evaluation. I touch on the risks of this work and related viral discovery efforts in Chapter 4.

Gene drives are another scientific development that have pushed the debate around misuse risks in recent years. Gene drives are genetic engineering constructs, based for instance on CRISPR/Cas9, which propagate a particular (modified) gene throughout a population. In the future, gene drives may be used to reduce the fertility or disease-carrying capacity of infectious disease vectors, including of malaria-transmitting mosquitoes (Esvelt et al. 2014). Notably, gene drives, as they do not resemble classical pathogens or toxins, were not covered by any existing oversight, despite their potential for misuse (Oye et al. 2014). Debates only touched on dual-use risks in passing, rather focussing on genetic engineering regulation and the transparency of experiments. Kevin Esvelt has proposed that the implications of gene drive research might be an opportunity to drive science to greater openness (Esvelt 2016). The release of self-replicating agents that change the environment will require great public trust; thus, researchers would need to engage local communities early in the research lifecycle to proactively address concerns. As self-spreading agents with significant environmental impacts that do not resemble classical organisms, gene drives are a prominent example of an innovation that challenges existing regulatory paradigms (S. W. Evans and Palmer 2018). In response to a 2016 National Academies Report, leading funding bodies of gene drive research have signed onto principles for responsible sponsorship (Emerson et al. 2017). Gene drives highlight how relevant stakeholders need to continuously revisit approaches to evaluate safety, security, and ethics of life sciences (Lunshof and Birnbaum 2017). For instance, current biosafety practices need to shift beyond a sole focus on high-risk human pathogens to also consider gene drive-like constructs.

Recently, a paper highlighted the dual-use potential of computational tools for the life sciences. An algorithm originally designed to find drug candidates with low toxicity to humans was able to similarly predict substances with maximal toxicity to humans, including agents previously used as chemical weapons and other not yet known potentially harmful chemicals (Urbina et al. 2022). The public attention on the biosecurity risks associated with artificial intelligence has only increased with the release of large

language models like chatGPT/GPT-4. During a July 2023 US Senate testimony, Anthropic Co-Founder and CEO Dario Amodei highlighted that large language models could enable the release of biological weapons as a critical medium term risk of artificial intelligence (Amodei 2023). I discuss the potential impacts of AI and associated dual-use risks in Chapter 7.

The historical biological weapon programs, bioterror incidents, and examples of dual-use research described in this chapter have dominated global debates on the topic of dual-use management. Thus, this history shapes existing national regulations and conceptions about misuse risks in the scientific community and the general population. Additionally, historical incidents have triggered various academic discussions, including in Science and Technology Studies, security studies, and bioethics. I will analyse this societal reception of the dual-use problem in the next chapter.

Chapter 2: Dual-use risks and society

Due to the subjective nature of assessing and weighing risks, societal factors have influenced the mitigation of dual-use risks. In this chapter, I assess conceptions about dual-use research across different areas of society. First, I give an overview of global dual-use regulations and relevant discussions in international fora (2.1). Then, I present how the broader scientific community has interfaced with dual-use risks (2.2). Lastly, I summarise sociological (2.3) and ethical analyses (2.4) of dual-use risks with particular relevance to my thesis.

2.1 Global regulations

Despite the historical use of biological weapons and debates about misuse risks, few countries regulate dual-use research explicitly. According to the 2021 Global Health Security Index, only 6% of countries have dual-use oversight for research involving especially dangerous pathogens. These countries are Armenia, Australia, Bulgaria, Canada, Slovenia, Brazil, Denmark, the United Kingdom, the United States, Qatar, Sweden, and Thailand (Nuclear Threat Initiative and Johns Hopkins Center for Health Security 2021). Germany and China also feature some relevant oversight mechanisms or budding regulations; these countries however did not qualify as featuring dual-use oversight under the criteria of the 2021 Global Health Security Index. Among countries with some form of dual-use oversight, the breadth of oversight differs. For instance, the United States oversees a moderate range of activities with its select agent-focussed DURC policy and function-based P3CO policy. In contrast, Canada's decentralised oversight, in theory, applies to a broader conception of dual-use research (Jacobsen et al. 2014). In the following, I discuss examples of regulations of different countries and the impact of dual-use debates globally.

2.1.1 United States

While not necessarily having the broadest reach, the United States' dual-use policies have been very influential globally and are frequently referenced. The three most relevant policies are the Federal Select Agent Program (FSAP), the DURC policies, and the P3CO policy. The FSAP aims to ensure laboratory security, which captures the physical security of pathogen and toxin specimens. It was initially created in 1996 after a microbiologist tried to purchase Yersinia pestis. The FSAP has been updated over time, with its last major update in 2012 when a subset of most concerning "Tier 1" agents was established (Centers for Disease Control and Prevention and U. S. Department of Agriculture 2022). FSAP oversees the possession of dangerous materials, approves registrations, performs inspections, and receives reports of select agent theft or loss. Under FSAP, the Federal Bureau of Investigation conducts security risk assessments of individuals working with select agents. The toxins and agents on the FSAP list are subject to export controls. The FSAP formed the context for the DURC policies of 2012 and 2014, which I introduced in the previous chapter. These policies form the core of dual-use management in the United States. The DURC policies require federal funders and federally-funded institutions to review the dual-use risk of research involving seven categories of experiments performed on 15 select agents and toxins, a subset of the FSAP list (see Table 1.1 and 1.2). I analyse DURC oversight and its limitations in detail in Chapter 8. Lastly, the P3CO policy of 2017 (discussed in the previous chapter) requires risk assessment and mitigation of HHS-funded studies involving the enhancement of potential pandemic pathogens.

2.1.2 European Union

In response to the H5N1 influenza studies, the European Academies Science Advisory Council (EASAC) set up a working group on GOF research. The group published its recommendations in 2015 (Fears and ter Meulen 2015). EASAC found that good practice was already in place at the member state level, where GOF research has to be justified to a range of bodies, from safety and ethics officers at the local institution

to funding bodies and national authorities (EASAC 2015). Thus, EASAC concluded that there was no need for a new advisory body at the EU level. Rather, EASAC advised that all EU Member States should instate a clear national mechanism for managing biosafety and biosecurity risks.

For EU-funded research, the Ethics Appraisal Scheme asks researchers to consider risks of misuse and features special requirements for relevant projects (European Commission 2022). The European Commission has issued guidance to appoint an ethics or security advisory board for relevant research to facilitate compliance with international, EU, and national regulations (World Health Organisation 2022b).

2.1.3 Germany

Since 1990, the Genetic Engineering Act has regulated genetic modification experiments in Germany (Federal Office of Consumer Protection and Food Safety 2022). This legislation requires institutions performing relevant experiments to have a biosafety officer who needs to attend training courses. Principal investigators are required to educate laboratory staff on biosafety and are personally liable for transgressions. Certain experiments need to be authorised by local authorities, and extensive records need to be kept (World Health Organisation 2022b). All research involving genetic modification of high-consequence pathogens needs to be reviewed by the Central Committee on Biological Safety (ZKBS).

Germany also considered dual-use risks more explicitly in the wake of the H5N1 avian influenza studies. In 2014, the German Ethics Council released a report featuring a series of recommendations. These included increasing education on dual-use risks, establishing a code of conduct, and reviewing federally funded research (German Ethics Council 2014; National Academies of Sciences, Engineering, and

Medicine 2016). The German government did not pass any formal policies in response to this report, however more informal measures were taken.

Since 2015, the German Research Foundation (DFG) and Germany National Academy of Sciences Leopoldina, through the Joint Committee on the Handling of Security-Relevant Research, have encouraged institutions to create dual-use oversight committees; many German institutions now have established such committees (Nationale Akademie der Wissenschaften Leopoldina 2022). Furthermore, the DFG conducts dual-use oversight for funded research.

2.1.4 United Kingdom

In 2005, the UK Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council (BBSRC), and the Wellcome Trust issued a joint statement on dual-use risks, which has since been updated (German Ethics Council 2014; BBSRC, MRC, and Wellcome 2015). This collaboration now includes a mechanism to allow concerns to be flagged and considered. The United Kingdom controls specific pathogens under the Anti-Terrorism, Crime and Security Act ATCSA of 2001. Relevant organisms are found on the Schedule 5 list, and responsible access is overseen by the National Counter Terrorism Security Office (Government of the United Kingdom 2001). Facilities looking to hold Schedule 5 substances have to register and receive regular visits by a Counter Terrorism Security Advisor from local police. Institutions engaging in genetic manipulation research also have to register with the Health and Safety Executive and need to establish an internal committee to review associated risks (Health and Safety Executive 2021).

2.1.5 Canada

Canada takes a comprehensive, decentralised approach to dual-use management. The basis of this is the Human Pathogens and Toxins Act, which was passed in 2009 and is reviewed and updated every five years (Government of Canada 2014). This act requires any organisation doing life sciences research, regardless of publicly or privately funded, to engage in dual-use oversight. Organisations need to create a plan for dual-use oversight and have this plan approved by the Public Health Agency Canada. Anyone working on a list of particularly concerning pathogens needs to undergo security clearance unless exempted and pathogen enhancement experiments need to be reported (Jacobsen et al. 2014).

2.1.6 China

In October 2020, the People's Republic of China passed a new biosafety and biosecurity law. The law does not include specific rules on GOF research but assigns responsibility for creating dual-use guidelines (Huang 2021). China might be especially important for future biotechnology governance, as it is becoming a key global player in biotechnology. This development exemplifies the importance of looking to international approaches to the governance of dual-use research.

2.1.7 United Nations

The United Nations (UN) interfaces with dual-use management in multiple ways. The UN administers the Secretary General's Mechanism for investigating biological and chemical attacks. United Nations Security Council Resolution 1540 explicitly prohibits nation-states to aid non-state actors in the acquisition of biological and chemical weapons (United Nations Office for Disarmament Affairs 2022). Lastly, the Biological Weapons Convention (BWC) is very relevant for dual-use management. An arms control treaty in force since 1975, the BWC has an Implementation Support Unit of three full-time staff housed in the United Nations Office of Disarmament Affairs (UNODA). Multiple articles of the BWC

relate to dual-use risks. The core of the BWC is Article I which forbids the development, acquisition, stockpiling, or use of biological substances with no peaceful or protective application (UNODA 2022). This broad ban is known as the general purpose criterion and has contributed to strong norms against biological weapons. Most relevant to dual-use research is Article III, which bans assistance to others in their efforts to develop biological weapons. Furthermore, Article IV obliges national implementation of the BWC (UNODA 2022); however, only 62% of signatories have banned the possession of biological weapons by non-state actors (Drobysz 2020). Article X protects the exchange of biological materials for peaceful purposes; it is often leveraged for international development goals and may provide a barrier to expanding dual-use management under the BWC. Despite previous attempts, no formal BWC verification mechanism exists (Lentzos 2019). In 1986, countries agreed to submit annual confidence-building measures (CBMs) on national peaceful activities, a commitment to which less than 50% adhere (Lentzos 2019). One promising proposal to strengthen dual-use management through the BWC are the Tianjin guidelines, a code of conduct for life scientists put forward in 2021 (Xue, Shang, and Zhang 2021). This proposal would expand the BWC's norm-setting role to dual-use research.

2.1.8 World Health Organisation

In 2002, the World Health Assembly highlighted the threat to public health posed by deliberate biological events (World Health Assembly 2002). The 2005 International Health Regulations and Joint External Evaluations touch on dual-use risks (World Health Organisation 2022a). In 2010, the WHO published guidance on responsible life sciences research (World Health Organisation 2010). Throughout the debates on the H5N1 studies, the WHO promoted national governance and review of dual-use research (World Health Organisation 2012). In September 2022, the WHO released a new "Global guidance framework for the responsible conduct of life sciences" (World Health Organisation 2022b). This framework provides a summary of existing approaches for dual-use management and sets out guiding steps to creating

dual-use oversight for different stakeholders. Notably, this framework adopts a quite broad view of dual-use risks; the presented case studies are not limited to the modification of high-consequence pathogens but also include work on viral vectors, gene drive, and neurological bioregulators.

2.1.9 The Australia Group

As part of the Australia Group, many countries coordinate export controls on a list of dangerous biological agents (The Australia Group 2020). This forum was established in 1985, and members include the United States, the United Kingdom, Australia, and many European Countries.

2.2 Perspectives in the scientific community

Scientists have for many decades thought about the societal implications of genetic engineering and other areas of the life sciences. The 1975 Asilomar conference established guidelines for recombinant DNA research and has been hailed as a great success of scientist-led research governance (Berg et al. 1975; Falkow 2012). I discuss Asilomar and more recent examples of scientific self-governance in the context of synthetic biology in Chapter 3, section 3.4. In the United States, discussions about the openness of life sciences research more generally go back to the Cold War. In 1982, the National Academies of Sciences published the Corson Report, which argued to keep science open despite its potential to inform the Soviet Union's advances (National Academies of Sciences, Engineering, and Medicine 1982). The goal was to achieve "security by accomplishment" rather than "security by secrecy". In 1985, the National Security Decision Directive 189 (NSDD189) mandated that the products of fundamental research should be unrestricted (National Security Decision Directives 1985). Discussions around the openness of scientific research flared up again after the discovery that Al-Qaeda had drawn on public and private scientific information for their attempts to attain biological weapons (Petro and Relman 2003). These discussions channelled energy into

the Fink report and eventual DURC policies. In 2011, the NSABB investigated ways to suppress the publication of the Kawaoka and Fouchier studies while nevertheless distributing public health-relevant information relating to the enhancement of H5N1 influenza (as discussed in section 1.3.1). However, the NSABB found no way to achieve this goal, and thus the work was eventually published largely in full. However, the debate on balancing openness and security of research output continues (DeFrancesco 2021; Smith and Sandbrink 2022).

Specific subsets of the scientific community are well aware of dual-use risks and think about them beyond the narrow scope of the US policies. In 2009, the American Association for the Advancement of Sciences polled its members on their attitudes toward dual-use risks (National Research Council 2009). Out of all the life scientists reporting work on dual-use research, only one in three noted working on one of the seven experiments of concern identified in the Fink report. This suggests that scientists have a broader conception of the dual-use potential of a research than just microbial threats. In 2006, the National Research Council published the report "Globalization, Biosecurity, and the Future of Life Sciences", which had discussed a broader range of misuse risks (National Research Council 2006). Despite the poll and this report, the existing DURC and P3CO regulations focus on select microbial threats. However, there may be a culture of voluntary compliance beyond these policies. According to Carrie Wolinetz, non-federally funded institutions voluntarily comply with DURC oversight (S. W. Evans et al. 2021).

Education is crucial for the effective governance of dual-use risks. The Fink report had already ended with the recommendation to educate the scientific community about dual-use risks (National Research Council 2004). However, education and awareness of dual-use risks have been limited (Minehata et al. 2013). Recent surveys find this remains the case. Only 41% of students competing in the International Genetically Engineered Machines (iGEM) competition knew the term "dual-use research" (Vinke, Rais,

and Millett 2022). In a survey in Pakistan, 58% of post-graduate researchers reported having heard the term "dual-use research of concern" (Sarwar et al. 2019). For improving dual-use management, improving dual-use education is key. This is especially true as the scientific community has called for retaining self-governance on dual-use issues (Resnik 2013b).

Different factors may make scientists less attuned to research risks. The nuclear scientist Oppenheimer highlighted how the evaluation of societal implications comes second to the drive for discovery:

> *"When you see something that is technically sweet, you go ahead and do it and you argue about what to do about it only after you have had your technical success."*
>
> - Oppenheimer 1954 (U. S. Atomic Energy Commission 1954)

Additionally, certain career incentives may go against careful consideration of dual-use risks. For instance, pressure to publish highly cited work means that researchers might perform particularly newsworthy and risky experiments. Lastly, Kitcher describes a commonly held belief that all knowledge is good in-and-of itself (Kitcher 2003). If such a belief exists, it may at least partially be due to a lack of exposure to biosecurity education. The danger of knowledge has long been acknowledged by the nuclear physics community, a community marked by the events of Hiroshima and Nagasaki (N. G. Evans 2013; Lanouette 1992).

One challenge for improving dual-use management - and likely scientific awareness - are disagreements about the likelihood of misuse. In a 2015 Delphi study, Boddie *et al.* asked 59 US biosecurity experts to estimate the probability of a large-scale bioterror attack over the subsequent ten years; opinions widely differed, ranging from 1% to 100%, with a median estimate of 57% (Boddie et al. 2015). Survey participants also disagreed about the level of sophistication and background of possible perpetrators.

The scientific community needs to engage more with the public on dual-use risks. In a survey which excluded individuals connected to life sciences research, 77% of participants were unaware that DURC is regularly conducted (MacIntyre et al. 2020). 64% deemed it unacceptable or said they were unsure about its acceptability; interestingly, providing more information on this research decreased acceptance. In the end, the results of dual-use research benefit or harm broader society; thus, ethical progress on dual-use management requires consideration of the opinions of an educated public.

2.3 Sociological perspectives

Sociology is crucial to understanding how the scientific community operates around research risks, how human factors play into misuse risks, and how risk mitigation strategies can leverage these factors. In the following, I present a subset of influential framings relevant to my arguments, drawing across the security studies and Science and Technology Studies (STS) literature.

2.3.1 Science in a social context

Kathleen Vogel presents a "biosocial frame" on dual-use risks in "Phantom Menace or Looming Danger?" (Vogel 2012). This frame aims to put science into its social context and thus consolidates a lot of previous work by STS scholars. Vogel's biosocial frame has four main components. First, it highlights the importance of know-how and other uncodified knowledge in the life sciences, which is an important barrier to misuse. Such uncodified knowledge has become known as "tacit knowledge", a description first proposed by Polanyi and refined by Collins (Polanyi 1974; H. Collins 2010). Second, the biosocial frame highlights the local, contingent character of biotechnology development and use. It argues that local expertise matters greatly for what and how experiments are conducted. Furthermore, scientific publications do not constitute how-to-manuals. Referencing STS scholar Bruno Latour, Vogel argues scientific publications do not describe what went on in the laboratory but rather serve as a summary of

relevant findings (Vogel 2012, 62; Latour and Woolgar 1986). Vogel highlights an extensive need for troubleshooting, which black-boxing descriptions of science frequently miss (Vogel 2012, 63). This need for troubleshooting is even present for popular and abundant Polymerase Chain Reaction (PCR) kits (Lynch 2002). Furthermore, large research projects are very complex and thus require significant managerial skills. These socio-organisational challenges have disrupted efforts to weaponise biology, including those of Aum Shinrikyo[5] (Leitenberg 2005; Danzig et al. 2012; Ouagrham-Gormley 2014).

The third and fourth components of Vogel's biosocial frame are particularly relevant to my thesis. The third component is the importance of understanding past work. Vogel argues for an incremental change-based model of progress rather than one marked by revolutionary jumps (Vogel 2012, 66). This has important implications for dual-use risks. If capabilities advance more incrementally, broader scientific developments would drive misuse risks rather than a small number of high-risk studies. Thus, a sole focus on DURC as a small subset of dual-use research may fail to mitigate research risks, an argument I develop in Part III.

The last component of the biosocial frame is the possibility of multiple biotechnology trajectories. The path of technology development is not fixed or technologically deterministic: rather there is room for human agency and societal factors to shape which trajectory is realised. This framing highlights the possibility of going beyond the oversight of individual projects and leveraging a broader approach for risk reduction, which I propose in Part IV.

[5] Discussed in Chapter 1, section 1.1.2.

2.3.2 Responsive science

Decision-making on potentially risky research needs to consider inputs of those potentially affected by the consequences. Sam Evans argues that science is currently "apart from society, but needs to become a part of society"[6] (S. W. Evans 2022). Science needs to serve society, be accountable, and respond to societal inputs.

Gene drives are one instance of biotechnology where the need for broader societal input is particularly apparent. Gene drives mediate self-propagating genetic changes. This may have very beneficial applications, like eliminating malaria; however, such applications may also have unpredictable effects on ecosystems or may violate local cultural beliefs (Oye et al. 2014). Thus, Kevin Esvelt, one of the inventors of CRISPR-based gene drives, has engaged extensively with local communities. He has proposed the concept of responsive science, a science responsive to societal inputs (Esvelt 2016). Relevant practices are the pre-registration of research before its initiation and and inviting public scrutiny of grant proposals; these practices would allow more time for a broader set of discussions about how work might best be governed (Esvelt 2017; 2016).

One challenge for societal inputs is the accelerating pace of biotechnology democratisation. Jackson et al. found that in 2019 novel biotechnologies took 4.5 years to become accessible to many individuals; by the end of 2030 this may decrease to 3.5 years (S. S. Jackson et al. 2019). Thus, slowing the dissemination of novel capabilities may be a useful governance intervention (Pannu et al. 2021); however, this would also be associated with costs and risks that need to be considered.

[6] Sam Weiss Evans used this exact wording in personal communication with me.

2.3.3 A theory of mitigating misuse

Different strategies for mitigating biotechnology misuse have been highlighted over the years. One such conceptualisation is the Swiss Cheese Model, which argues that prevention is not about preventing any single step but rather a combination of risk mitigation approaches (Perneger 2005). Even if a given risk mitigation strategy does not fully prevent misuse, it raises the bar and makes it more costly, and thus less likely. For instance, the BWC is currently not enforced; nevertheless, it has created strong norms against biological weapons, which force relevant research into secrecy and thus increase its cost. Sandberg and Nelson have modelled that for successful misuse that is dependent on a series of steps of unequal difficulty, it is the most difficult step that limits most actors (Sandberg and Nelson 2020). Thus, Sandberg and Nelson conclude that highly skilled/powered actors are most likely to succeed at misuse, even if less skilled actors with malicious intent are much more abundant.

STS scholar Sheila Jasanoff has highlighted the importance of humility when conducting and regulating science. It is difficult to predict outcomes and societal impacts of technologies and regulations. Thus, Jasanoff calls to develop technologies of humility, methods to make decisions about technological progress given uncertainty (Jasanoff 2007). In this vein, more recently, a large number of biosecurity experts have called for experimentation around the assessment and mitigation of dual-use risks (S. W. Evans et al. 2020). As dual-use risks become more abundant, new approaches and frameworks need to be tested, a process I hope to contribute to through this thesis.

2.4 Ethics of dual use

Ethicists have discussed different aspects of dual-use research, technology, and knowledge. As well as commenting on specific dual-use experiments, ethicists have analysed the moral dimensions of dual-use research (Sture 2013; Forge 2013), the societal dimensions of making relevant decisions (S. Miller 2013;

N. G. Evans 2014), and norms against biological weapons (S. Miller and Selgelid 2007; van der Bruggen 2013). However, generally, ethicists have given relatively little attention to dual-use research, despite its relation to ethical debates on genetic engineering and its implications for catastrophic risk (Selgelid 2013). Aiming to bring dual-use considerations into mainstream bioethics, Thomas Douglas and Julian Savulescu have proposed to create an "ethics of knowledge" (Douglas and Savulescu 2010). While ethical consideration of dual-use risks spiked after the publication of the H5N1 studies - for instance, a comprehensive volume by Rappert and Selgelid was published in 2013 (Rappert and Selgelid 2013) - contemplation of dual-use research has waned since then. In the following, I present ethical commentary on the dual-use examples discussed in Chapter 1. I focus on analyses of decision-making on dual-use research to provide background for the arguments I develop in Parts III and IV.

Ethical commentary accompanied the various dual-use debates. In the wake of the IL-4 mousepox debate, Miller and Selgelid analysed the dual-use dilemma in the life sciences in 2007 (S. Miller and Selgelid 2007). They highlighted how the dual-use dilemma exists for both researchers and the government. In particular, Miller and Selgelid advocated for resolving the dual-use dilemma through a "third option" beyond conducting or not conducting the research; I expand on this idea in Part III and IV. Additionally, they analyse research risks beyond the experiments defined in the Fink report as experiments of concern, including discussing dual-use risks of sharing pathogen sequences. In their ethical analysis, Miller and Selgelid highlight that the ethics of dual-use touches on many values beyond human health; these values include the right to life, free speech, academic freedom, and justice. Miller and Selgelid highlight academic freedom as a crucial objection to dual-use oversight, citing Oppenheimer: *"Secrecy strikes at the very root of what science is, and what science is for."* (Schweber 2000). Some may see academic freedom as necessary for scientific progress. However, Miller and Selgelid conclude that academic freedom should be subordinate to security concerns. They argue that academic freedom and resulting scientific progress is not an all-or-

nothing question, as shown by the fact that nuclear physics has advanced despite security regulations. Nick Evans makes a similar argument around limitations on human experimentation and highlights that the space of research that may be funded is already much smaller than all possible research proposals (N. G. Evans 2018). Miller and Selgelid also consider a related practical objection by Ian Ramshaw, the author of the mousepox publication. Ramshaw argues that it is too late for censorship in the life sciences as sufficient information is already public for causing great harm. However, Miller and Selgelid argue that future publications might be worse, and additional publications might make the offensive uses of existing public knowledge more obvious.

In the wake of the gain-of-function debate, Selgelid published a second influential ethical analysis (Selgelid 2016). The US government commissioned this ethical analysis to consider ethical considerations of GOF research (GOFR). At the core of Selgelid's analysis lies the definition of concrete principles for deciding whether to fund and conduct an experiment (Box 2.1). Decision makers can evaluate experiments on these principles to identify their level of acceptability and possible risk mitigation measures. I apply Selgelid's principles to dual-use research beyond GOFR in Chapter 9 and in Appendix B.

Box 2.1: Selgelid's framework for gain-of-function research (GOFR) decision- and policy-making

Quoted from (Selgelid 2016)

1. *Research Imperative:* The ethical acceptability of GOFR posing extraordinary risks partly depends on the importance of the research question it aims to address.
2. *Proportionality:* The ethical acceptability of extraordinarily risky GOFR partly depends on the extent to which there is reasonable expectation that the research in question will (1) yield answers to the target public health question and (2) ultimately result in benefits that outweigh risks involved.
3. *Minimisation of Risks:* Other things being equal, the ethical acceptability of a GOFR study is a function of the degree to which (1) there is confidence that no less risky forms of research would be equally beneficial and (2) reasonable steps have been made to minimise risks of the GOFR study in question.
4. *Manageability of Risks:* Other things being equal, the more manageable the risks of a GOFR study, the more ethically acceptable the study would be. Conversely, the more important/beneficial a GOFR study is expected to be, the more we should be willing to accept potentially unmanageable risks.
5. *Justice:* Because justice requires fair sharing of benefits and burdens, the ethical acceptability of GOFR partly depends on the degree to which (1) risks fall on some people more than others, (2) risks fall on those who are unlikely to benefit, and/or (3) any resulting harms are uncompensated.
6. *Good Governance—Democracy:* GOFR decision- and policy-making should (insofar as possible) reflect the ultimate values, value weightings, and risk-taking strategies of public citizens.
7. *Evidence:* Decision- and policy-making regarding GOFR should be based on more/better evidence regarding risks, benefits, (means of) risk minimisation, who is likely to benefit or be harmed by research, and the values, value weightings, and risk-taking strategies of public citizens.
8. *International Outlook and Engagement:* Because risks and benefits of GOFR (can) affect the global community at large, the ethical acceptability of GOFR partly depends on the extent to which it is accepted internationally. Decision- and policy-making regarding GOFR should

> (insofar as possible) involve consultation, negotiation, coordination, and related forms of active engagement with other countries.

Others also conducted ethical analyses of the gain-of-function studies. The German Ethics Council published a report in 2014 (German Ethics Council 2014). David Resnik conducted an ethical analysis in which he particularly considered the ethical acceptability of risk mitigation strategies (Resnik 2013a). He concludes that redacted publication of the H5N1 studies would have been a reasonable and potentially favourable strategy. However, he acknowledges the practical and legal constraints which eventually led to full publication of the manuscripts.

2.4.1 Cost-benefit assessments

Weighing risks and benefits, in other words a cost-benefit approach, is the most frequently discussed strategy for making decisions about dual-use research. For instance, Tom Douglas has proposed a very explicit expected value approach (Douglas 2013). The most comprehensive attempt at a risk-benefit assessment of dual-use research is the study on GOF research contracted by the US government (Gryphon Scientific 2016). In this study, Gryphon Scientific concluded that quantifying the risks and benefits of gain of function research is very difficult, with estimating benefits being even more challenging than estimating risks.

A crucial component of cost-benefit assessments is considering alternative options. Solely assessing the risks and benefits of a single experiment may miss the fact that researchers could have used their funding and time for other experiments with potentially more favourable risk-benefit profiles. Selgelid's decision-making principles also incorporate the evaluation of alternative experiments. For instance, Selgelid's Minimisation of Risks principle states that a dual-use study is only acceptable if no less risky alternatives

are available (Selgelid 2016). Nick Evans points out that a significant failure of the US P3CO policy is that it fails "to consider that alternative experiments may provide a superior expected net benefit" (N. G. Evans 2018). All of these discussions still compare risky experiments to other studies yielding similar benefits, such as other studies that may help to assess future pandemic threats. However, as the range of research featuring significant dual-use risks grows and as certain lines of inquiry can be broadly considered dual-use, cost-benefit comparison across research fields becomes an important lever for reducing risks. I describe examples of the changing risk landscape in Chapters 4-7 and expand on the argument for considering a broader set of low-risk alternatives in Chapter 10.

2.4.2 The precautionary principle

The precautionary principle is another decision-making tool sometimes used in relation to dual-use research. The precautionary principle emerged in Germany and Sweden in the 1960s (Sunstein 2005); its development was partially motivated by shortcomings of cost-benefit analysis (Clarke 2013). Many different forms of the precautionary principle have been defined, some complementary to cost-benefit analysis (Clarke 2013). Steve Clarke highlights versions of the precautionary principle that can help address costs that are not fully quantifiable and assign a burden of proof for research risks (Clarke 2013).

Kuhlau *et al.* have formulated a precautionary principle for dual-use life sciences research that focuses on precautionary measures:

> *"When and where serious and credible concern exists that legitimately intended biological material, technology or knowledge in the life sciences pose threats of harm to human health and security, the scientific community is obliged to develop, implement, and adhere to precautionary measures to meet the concern."* (Kuhlau et al. 2011)

Clarke criticises this precautionary principle as being unclear in its goals. If conceptualised as a strong guiding principle and thus an alternative to cost-benefit analysis, the principle lacks clarity on how to consider differences in benefits across options and how to avoid a state of paralysis when the precautionary principle points against all available options (Clarke 2013).

Societal perspectives on dual-use risks discussed in this chapter define the existing paradigm to dual-use management, a paradigm which I analyse more explicitly in Chapter 8 and critique throughout my thesis. Regulations and decision-making strategies for dual-use research need to evolve with new risks and a changing scientific landscape. Thus, I turn in the next chapter to advances in synthetic virology, which underpin emerging dual-use risks and appropriate responses discussed throughout my thesis.

Chapter 3: Advances in synthetic virology

Scientific advances shape dual-use risks. As new capabilities have emerged over the years, such as the reconstruction of viruses from synthetic DNA, dual-use debates have followed. In this chapter, I provide an overview of scientific advances that define existing capabilities. These scientific developments may also inform where future dual-use risks will emerge. In particular, I evaluate the advances in synthetic biology and, more concretely, synthetic virology. I focus on research relating to viruses, as viruses pose particularly great catastrophic risks. Viruses can evolve rapidly, may be highly transmissible, and are not easily countered by broad-spectrum therapeutics.

Many scientific advances with impacts on dual-use risks can be traced back to the emergence of synthetic biology. I use "synthetic biology" to refer to a set of new practices for creating and modifying biological organisms. More specific definitions of synthetic biology are contended (Nature Biotechnology 2009). Synthetic biology has been described from different perspectives since the early 2000s (Benner and Sismour 2005; Endy 2005; Dhar and Weiss 2007). All of these perspectives share an anticipation for synthetic biology to revolutionise biological engineering. Synthetic biology shifts the goal of studying biology from learning about life to the application of biology as a technology. This perspective has been reinforced by Thomas Dixon and colleagues who argue that core assumptions of synthetic biology are that life is information and biology is technology (Dixon et al. 2022).

Synthetic virology describes new forms of re-creating and modifying viruses, and thus may be seen as a subfield of synthetic biology. Generally, synthetic virology is used to study viruses and design novel therapeutics (DeFrancesco 2021). Synthetic virology relies on reverse genetics approaches, genetic approaches to create a virus with a desired or modified function from genetic material. Necessary capabilities for streamlined reverse genetics include DNA synthesis, assembling small DNA fragments into

a viral genome, and rescuing or "booting" a virus from genetic material like DNA or RNA. The ability to synthesise viruses in this way has enabled a fundamentally different way to attain viruses; previously, access was dependent on acquiring samples from patients, nature, or other laboratories. Thus, synthetic virology greatly impacts misuse risks. Aum Shinrikyo considered harvesting Ebola virus from patients in 1992 (Danzig et al. 2012). Aum Shinrikyo's lead scientist, Seiichi Endo, studied virology at Kyoto University. Given Endo's relevant training, today, Aum Shinrikyo might have been able to create Ebola virus from publicly available genetic information with a limited number of equipment and materials. Thus, advances in synthetic virology significantly contribute to the shift in dual-use risks described by Lewis and colleagues: increasingly, insights and knowledge than physical pathogen samples exhibit the greater share of dual-use risks (Lewis et al. 2019).

This chapter aims to provide an overview of scientific advances that shape emerging risks. I first discuss how and what capabilities have emerged (3.1) and their benefits (3.2), before analysing the democratisation of synthetic biology (3.3) and associated governance efforts (3.4).

3.1 The evolution of the modern viral engineering pipeline

Synthetic biology has its roots in the field of genetic engineering; in 1974 Walter Szybalski already used the term synthetic biology and foresaw capabilities that now exist (Szybalski 1974). Synthetic biology emerged properly almost 30 years later. Two feats published in the year 2000 are frequently described as the origins of synthetic biology. First, Elowitz and Leibler successfully created an oscillating network of transcriptional regulators akin to a synthetic biological clock (Elowitz and Leibler 2000). Second, Gardner and colleagues built genetic switches which could be activated through chemicals or heat (Gardner, Cantor, and Collins 2000). Both of these publications had in common that they leveraged approaches from engineering. In particular, they used a design-build-test cycle, which has since become a hallmark of

synthetic biology. To facilitate the use of biology as building blocks for technology, standardised genetic parts and DNA regulators are nowadays available, alongside tools to simulate biological designs and open source DNA assembly (Knight 2003; Galdzicki et al. 2011; Gibson et al. 2009; National Academies of Sciences, Engineering, and Medicine 2018).

Increasingly accessible methods to synthesise viruses from synthetic DNA are particularly important. These capabilities go back to 2002, when Wimmer and colleagues recreated poliovirus (Cello, Paul, and Wimmer 2002). As previously mentioned, in 2005, researchers replicated this feat for one of the most deadly pathogens in history, the H1N1 influenza virus which caused the 1918 pandemic (Tumpey et al. 2005). Synthetic biology capabilities spread very rapidly as anyone with relatively basic skills, materials, and equipment can replicate methods shared online. While the work on the initial reconstruction of poliovirus of 2002 took three years, just two years later, the J. Craig Venter Institute synthesised a similarly sized virus in just two weeks (G. Gronvall 2016, 38).

Another impactful advance is easy and precise gene editing. In 2012, Emmanuelle Charpentier and Jennifer Doudna discovered the possibility of leveraging the bacterial immune system components CRISPR/Cas9 for making precise cuts in DNA (Jinek et al. 2012). Since then, CRISPR/Cas 9 genome editing has rapidly spread in its application; within a single year it was extended from cell culture in the US to being used to create genetically modified monkeys in China (Mali et al. 2013; Niu et al. 2014). CRISPR/Cas9 gene editing has significant security implications (National Academies of Sciences, Engineering, and Medicine 2018). However, compared to other organisms like bacteria, the creation of engineered viruses is less dependent on gene editing because of the possibility of synthesising genetically modified viruses from scratch. Such *de novo* synthesis of viruses is possible due to their relatively small and simple genomes. The synthesis of larger viruses, such as poxviruses, is still relatively difficult. Thus, gene

editing approaches play a more significant role in engineering these larger viruses (Boutin, Mosca, and Iseni 2022).

3.1.1 The viral synthesis pipeline

Over the last two decades, the synthesis of viruses has become cheaper and more accessible. There is a great incentive to make viral synthesis even easier, faster, and cheaper as this can speed up the iteration of design-build-test cycles for novel therapeutics and other applications of synthetic viruses. Improvements across its core steps advance viral synthesis capabilities: the design of synthetic genomes, DNA synthesis, DNA assembly, and viral booting (see Figure 3.1).

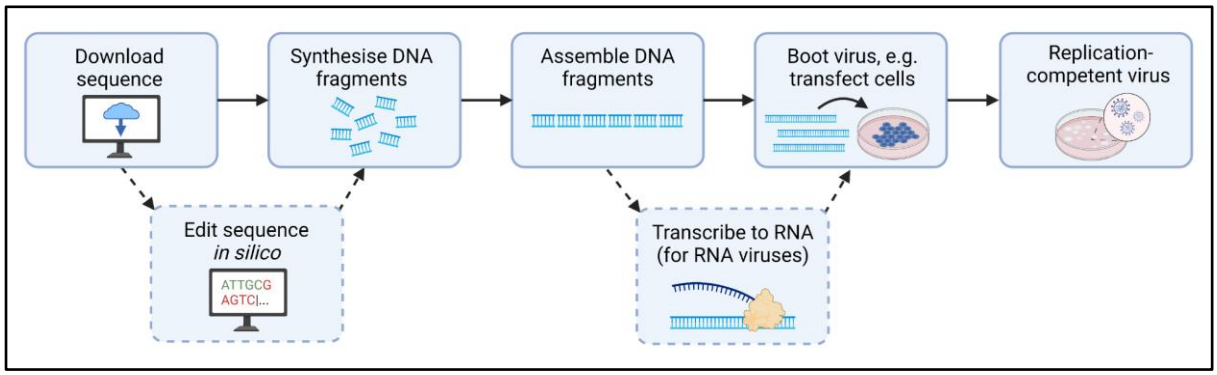


Figure 3.1: The modern viral synthesis pipeline

Viral genomes can be downloaded and optionally edited in a text editor or specialised program. DNA fragments can be ordered online and then assembled into full-length genomes. For RNA viruses, the genome needs to be transcribed into RNA. Booting a virus from its genome may take multiple forms, for instance transfection of cells.

Viral genomes. The first step for viral synthesis is downloading and potentially editing its genetic blueprint. Large databases of genomes and genetic components are publicly available on platforms like GenBank or GISAID, including specific platforms for viral genomes (Pickett et al. 2012). Open source software exists that can guide genome modifications (DeFrancesco 2021). Once the genome of a target virus has been curated, the next step is DNA synthesis.

DNA synthesis. Cheap DNA synthesis has been crucial for advancing synthetic virology. Over the past decades, synthesis costs have fallen and thus do not present a substantial barrier for viral synthesis (Hughes and Ellington 2017; Kosuri and Church 2014). The most commonly used form of synthetic DNA are DNA fragments of less than 100bp, so-called oligonucleotides. Oligonucleotides are now available for under 10 cents per base pair (bp).[7] These gene fragments still require assembly to form whole genomes. Nowadays, ordering assembled gene fragments of 5-10 kilo bases (kb) or longer is also possible, as gene synthesis companies may assemble smaller fragments before shipment. However, this may be associated with additional scrutiny and customer screening.[8] While researchers usually order synthetic DNA products from DNA synthesis companies, benchtop DNA synthesis devices are becoming increasingly powerful (see 3.4.4) (World Economic Forum and Nuclear Threat Initiative 2020).

Quality control and error reduction of DNA synthesis are also improving. This can be either through new techniques to reduce errors during gene synthesis (Ma, Saaem, and Tian 2012) or through the falling cost of DNA sequencing, the reading of DNA. In contrast to DNA synthesis, DNA sequencing costs have fallen much more steeply. While it cost $100m to sequence a human genome in 2001, it cost less than $1000 in 2021 (National Human Genome Research Institute 2021). This has sped up the quality control and test parts of the design-build-test cycle.

DNA assembly. DNA assembly is crucial to turning synthetic DNA fragments into complete viral genomes. New methods, such as the Gibson assembly, have increased the ease with which large genomes can be assembled (Gibson et al. 2009). Gibson and colleagues used the Gibson assembly method to create

[7] The cheapest relevant products on Twist's website price at around 7 cents/bp. (Twist 2023)

[8] According to an assessment by Kevin Esvelt.

the first synthetic bacterial genome in 2008 (Gibson et al. 2008). Polymerase cycling assembly modifications have enabled full-genome amplification, and the creation of copies of the whole viral genome (Bryksin and Matsumura 2010; Dean et al. 2001; DeFrancesco 2021). Due to advances in DNA synthesis and assembly, skilled researchers can nowadays create synthetic viral genomes in two to three weeks (DeFrancesco 2021). Despite these advances, it still requires a large amount of experience and skill to successfully create large viral genomes, such as genomes of poxviruses.

Converting DNA into RNA. This step is only necessary for viruses with an RNA genome. One example is poliovirus, a plus-stranded RNA virus. To create a poliovirus genome, a T7 RNA polymerase can turn complementary DNA templates into genomic RNA.

Viral booting. Viral booting or rescue describes the step of turning a synthetic genome into a functional virus. Traditionally, the technical difficulty of viral booting has been considered a significant hurdle for misuse (National Academies of Sciences, Engineering, and Medicine 2018). However, technical advances and detailed protocols have lowered this barrier - however, how much is contentious (Schulson 2022). The difficulty of booting a virus depends on the size of its genome and whether the genome is naturally infectious. Poliovirus is small and its genome is infectious, thus it is straightforward to boot it from its genomic RNA. Double-stranded DNA viruses like SV40 generally can be rescued by inserting their genetic materials into cells. Many viruses, including herpesviruses and poxviruses, require more complicated and specific viral rescue protocols, for example, the use of a helper virus (DeFrancesco 2021; Noyce, Lederman, and Evans 2018). The publication of detailed protocols is decreasing the expertise required to boot viruses with misuse potential. For instance, Xie *et al.* published step-by-step instructions for synthesising potentially engineered variants of SARS-CoV-2, which, together with colleagues, I have flagged as concerning from a dual-use perspective (Xie et al. 2021; Pannu et al. 2021).

3.1.2 Advances beyond viral synthesis

Beyond the viral engineering pipeline, other methods may increase viral engineering capabilities. One important method for optimising viruses for certain properties is to leverage evolution to optimise a virus for a given purpose. In directed evolution experiments, an organism is put into an environment to produce offspring, and then a subset of this offspring is selected for the desired function. One example of a directed evolution experiment is Fouchier's H5N1 study (see 1.3.1), where Fouchier and colleagues passed the virus through a series of ferrets and thus selected for mutations with mammalian transmissibility (Herfst et al. 2012). New directed evolution tools allow for optimising viruses across multiple properties and in a much greater parallelisation; machine learning tools can enhance this process (Ogden et al. 2019). Generally, artificial intelligence-empowered computational tools will play an increasing role in viral engineering, making capabilities previously reserved to a handful of experts accessible to many individuals (Jumper et al. 2021; Leman et al. 2020). I will discuss the impact of artificial intelligence in more detail in Chapter 7.

3.2 Benefits of synthetic biology and virology

Synthetic biology offers tremendous opportunities to improve the health and wealth of humanity and the environment. It drives a growing global bioeconomy, a sector estimated to have the potential to scale to $4-30 trillion USD (Hodgson, Maxon, and Alper 2022). According to 2018 estimates of US revenues, engineered plants and microbes already gross over $300 billion, and biotechnology-based industrial products yield annually over $115 billion (National Academies of Sciences, Engineering, and Medicine 2018).

Besides impressive revenue numbers, concrete advances highlight the impact of synthetic biology. The antimalarial artemisinin used to be derived from plants grown in China, Vietnam, and Kenya, a costly process associated with regular shortages (G. Gronvall 2016, 5–6). Now, artemisinin is synthesised in genetically modified yeast, which has decreased the costs and increased the global supply of this essential medication (Martin et al. 2003; Westfall et al. 2012). Synthetic biology has contributed significantly to the global response against COVID-19. Life-saving mRNA vaccines were designed on computers and produced using synthetic biology methods (Dixon et al. 2022). Synthetic biology-enabled experimental methods have contributed to the rapid characterisation of SARS-CoV-2 and informed the design of countermeasures. For instance, Jesse Bloom and colleagues have engineered yeast to display SARS-CoV-2 surface protein, enabling the comprehensive identification of possible antibody-evading mutations (Starr et al. 2020). Military applications of synthetic biology have also been explored. For instance, the US Department of Defense explored using synthetic biology to create environmentally friendlier explosives, which sparked debate about the ethics of such military applications (Hayden 2011).

Synthetic virology offers particular promise to enable novel vaccines and therapeutics. These products often leverage viral vectors, non-pathogenic synthetic viruses engineered to express a specific protein. For instance, the Oxford and Johnson & Johnson COVID-19 vaccines were based on viral vectors (Falsey et al. 2021; Sadoff et al. 2021). Viral vector-based gene therapy may enable the treatment of previously incurable heritable diseases (Shahryari et al. 2019). Several such gene therapy products have now been licensed, including a product for certain inherited retinal defects (Dias et al. 2018). Virus-based therapeutics are a particularly promising area of cancer research. Viral immunotherapies induce anti-cancer immune responses, while so-called oncolytic viruses are engineered to directly target and kill cancer cells. Over a hundred clinical trials of oncolytic viruses are registered on ClinicalTrials.gov.[9] In 2015, an

[9] On 19 August 2022, the search term "oncolytic virus" on ClinicalTrials.gov generated 145 registered trials.

engineered herpesvirus for treating melanoma became the first licenced oncolytic drug in the US and Europe (Rehman et al. 2016). While synthetic virology-enabled therapeutics feature potentially significant upsides, associated research can have dual-use potential. I analyse the dual-use risks of viral vector vaccines in Chapter 5 and viral vector-based gene therapies in Chapter 6.

Lowering the barrier to acquiring and researching viruses generally benefits research. Nowadays, researchers can synthesise a virus rather than having to engage in the paperwork-intensive process of accessing patient samples or sample banks. Indeed, Rourke *et al.* suggest that one motivation for David Evan's synthesis of horsepox virus was to circumvent Mongolian authorities (Rourke, Phelan, and Lawson 2020). However, Rourke *et al.* highlight another motivation for Evans and colleagues might have been avoiding the need to share the benefits of their research. Synthetic reconstruction of viruses based on public blueprints may be used to commercialise organisms without giving back to the local communities in which they were identified.

Lastly, synthetic biology has the potential to be a solution for reducing research risks. Researchers may modify organisms to prevent their replication outside of a controlled environment, a practice termed intrinsic biocontainment. For example, organisms may be engineered to depend on a constant level of small molecules, like estradiol, in their surroundings (Cai et al. 2015). A different approach may be to redesign crucial viral proteins to depend on non-natural amino acids, molecular building blocks that are not found in hosts outside of a controlled laboratory environment (Mandell et al. 2015).

3.3 The democratisation of bioengineering

Arguably, advances in synthetic biology are leading to a progressing democratisation of bioengineering. New methods to create viruses from synthetic DNA have lowered the barrier to the acquisition and misuse

of potential pathogens. Thus, an increasing number of individuals can misuse dual-use information, such as genetic blueprints, step-by-step experimental protocols, and specific insights on pathogen enhancement. Specific developments relating to democratised bioengineering deserve special attention.

3.3.1 Accessibility

High schoolers can now perform genetic engineering feats that 10 years ago were reserved for cutting-edge scientists. The synthetic biology competition iGEM showcases this, a competition in which every year 6000 high schoolers and undergraduates compete in 300 teams from over 40 countries (Millett et al. 2019). Since its creation in 2004, iGEM has produced over 30,000 individuals with basic synthetic biology skills. Many teams perform cutting-edge research, creating synthetic organisms with applications of substantial benefit. For instance, one team from Imperial College, London, created a synthetic bacterium that produces cellulose, which may find commercial application in water filters (Florea et al. 2016; G. Gronvall 2016, 11–12).

3.3.2 DIY biology

Do-It-Yourself (DIY) biology is a new community that engages in synthetic biology outside of traditional institutions. This community has arisen from the intersection of synthetic biology and "hacker- and makerspaces" (Sundaram 2021). In particular, the DIY biology community has made news through "biohacking"; for instance, the Food and Drug Administration (FDA) has reported concerns about the self-administration of DNA constructs (Mullin 2017). The lack of clear lists of laboratories and their activities challenge traditional safety and security oversight. However, community laboratories are still frequently subject to local regulations; for instance, in the UK, genetic modification experiments have to be declared to the Health and Safety Executive (Sundaram 2021). Despite the convention-defying approach of the DIY community, many academics judge that the community is taking safety and security

education seriously (Kuiken 2016; G. K. Gronvall 2018; Jefferson, Lentzos, and Marris 2014). For example, the DIY community worked together with leading biosafety experts to create a community biosafety handbook (Armendariz et al. 2020; Sundaram 2021). The rise of DIY biology means that new policies have to consider research spaces outside of classical institutions and renews the need for fostering a culture of trust, accountability, and responsibility.

3.3.3 Biofoundries

Biofoundries, sometimes also called cloud laboratories, are automated facilities that produce synthetic organisms to order. These facilities combine advances in robotics and synthetic biology and thus can "generate hundreds or thousands of constructs/strains in just a few days" (Vickers and Freemont 2022). These biofoundries mean that soon someone hoping to use synthetic biology will no longer need technical skills or equipment; they might simply order an automated facility to do the work. This has important security implications, as malicious actors might leverage these biofoundries to generate agents of misuse. Appropriate security protocols may mitigate these risks.

3.3.4 Benchtop synthesis machines

Benchtop synthesis machines enable DNA synthesis directly in laboratories and remove the need to order from DNA synthesis companies. The first commercial benchtop devices are now available (DNA Script 2021; Kilobaser 2022). These devices will likely disseminate the ability to create synthetic DNA and thus will have significant security implications. As I will discuss in the next section, benchtop synthesis machines provide a challenge for DNA synthesis screening (World Economic Forum and Nuclear Threat Initiative 2020).

Advances relating to synthetic biology point towards a proliferation of misuse risks. However, Catherine Jefferson and colleagues have questioned whether advances in synthetic biology are actually lowering the barrier to the misuse of biological agents (Jefferson, Lentzos, and Marris 2014). While synthetic biology has brought standardised biological parts and new tools, DNA synthesis has not dramatically improved over the last decade, and the assembly of DNA fragments remains a technical challenge (Kosuri and Church 2014; Jefferson, Lentzos, and Marris 2014). Thus, Jefferson and colleagues argue that the main challenge to misuse remains: the tacit knowledge barrier and the troubleshooting of experiments.

3.4 Synthetic biology and governance

The history of genetic engineering is tightly interlinked with discussions of societal implications. In 1975, Paul Berg, Maxine Singer, and colleagues convened the Asilomar conference to preemptively address potential biohazards (Berg et al. 1975). The conference established guidelines for recombinant DNA research and is remembered as a role model of scientific self-governance (Falkow 2012; Making and Hanna 1991). The echo of Asilomar was heard during the H5N1 GOF debate and might have inspired the 2012 voluntary moratorium (Casadevall and Shenk 2012; Casadevall and Imperiale 2014).

Like genetic engineering, synthetic biology's emergence has also been intertwined with the consideration of security risks. Synthetic biologists discussed the security implications of their work at the first synthetic biology conferences of the early 2000s. Proposals for risk mitigation measures included gene synthesis screening and a hotline for reporting safety and security concerns (Maurer, Lucas, and Goldman 2006; G. Gronvall 2016, 36). This proactive stance on self-governance was supported by broader society; the Alfred P. Sloan Foundation has funded governance efforts and the FBI has supported outreach programs to raise awareness for misuse risks (G. Gronvall 2016, 37).

Synthetic biology has become a case study for emerging technology governance. For instance, Brett Edwards has used synthetic biology to illustrate the innovator's paradox, which describes that innovation can produce both positive and negative consequences. Edwards argues that the global nature of synthetic biology highlights the general importance of global innovation governance (Edwards 2019).

In 2018, the US National Academies of Sciences published a comprehensive report on the biosecurity implications of synthetic biology. The report found that the highest security risks are posed by the re-creation of known pathogenic viruses, new approaches for synthesising biochemicals, and enhancing existing bacteria. The enhancement of existing viruses fell into the second highest category of risk. The WHO and others have also identified these and other areas in horizon scans for emerging security risks (Wintle et al. 2017; Kemp et al. 2020; World Health Organisation 2021b). Throughout my thesis, I discuss risks from the re-creation of known pathogenic viruses and the enhancement of existing viruses; I highlight how new research approaches contribute to these risks in Part II and propose solutions for addressing these risks in Part III and IV.

3.4.1 DNA synthesis screening

One particularly promising governance intervention is DNA synthesis screening. In 2006, The Guardian reporter James Randerson ordered a small fragment of smallpox DNA, a feat that highlighted a significant security gap (Randerson 2006). Despite technical caveats around its presentation, this story pushed public awareness of the need to screen synthesis orders.

In 2009, a group of DNA synthesis companies formed the International Gene Synthesis Consortium (IGSC) and committed to DNA synthesis screening. As part of its harmonised screening protocol, IGSC

members check the identity of customers and screen DNA synthesis orders against databases of known pathogens (Diggans and Leproust 2019). For its regulated pathogens database, the IGSC draws on international lists of controlled pathogens of the US Federal Select Agent Program, US Commerce Control, the Australia Group, and the European Union (International Gene Synthesis Consortium 2017; DeFrancesco 2021). The IGSC currently only screens double-stranded DNA sequences of length greater than 200bp. The IGSC has since become a champion for biosecurity, actively engaging and shaping international discussions on DNA synthesis and related topics (World Economic Forum and Nuclear Threat Initiative 2020). Many DNA synthesis companies, especially smaller ones, are still not part of the IGSC; thus, a significant fraction of the DNA synthesis market does not yet screen any orders.

In 2010, the US government created DNA synthesis screening guidance (U.S. Department of Health and Human Services 2010). Like current IGSC practice, the US government recommended in this guidance to screen double-stranded DNA (dsDNA) sequences of greater than 200 bp. This US screening guidance is currently being revised (Federal Register 2020). Initial drafts of the new guidelines suggest an ambitious step-up; this includes the screening of oligonucleotide orders of less than 100bp and of nucleic acids beyond dsDNA. However, the proposed screening of oligonucleotide orders may push the limits of voluntary synthesis screening; flagged orders are investigated by experts, which is expensive. Thus, in a low-margin market like oligonucleotide synthesis, the new screening guidance might disadvantage companies engaging in screening (Administration for Strategic Preparedness & Response 2020, 41). Governments need to consider other ways to incentivise screening practices. The US states Maryland and California have previously explored mandating screened synthetic DNA products for government-funded research (DeFrancesco 2021). While both initiatives ultimately failed, a recent smaller bill is now mandating public universities in California to only use screened gene synthesis products (California State Government 2022).

Academics and non-profits are currently developing improved screening technology. The Nuclear Threat Initiative and World Economic Forum published in 2020 a report on advancing a common mechanism for DNA synthesis screening (World Economic Forum and Nuclear Threat Initiative 2020). This common mechanism aims to provide cost-effective and comprehensive screening technology; integration into benchtop devices is also being explored. Another new screening approach is SecureDNA, an initiative initiated by Kevin Esvelt at the Massachusetts Institute of Technology ('Secure DNA Project' 2022; Gretton et al. 2021). SecureDNA aims to provide fully automated and secure screening for DNA fragments as small as 30bp. SecureDNA considers that certain screening targets may constitute security-sensitive information and thus features a cryptographic solution to safeguard these sequences.

3.4.2 iGEM as a biosecurity champion

The iGEM competition has also recognised the security implications of synthetic biology. The competition has a safety and security team, which not only reduces risks across the competition but also has become a source of international advocacy for biosecurity (Millett et al. 2019). iGEM takes the education of competition participants on safety and security very seriously and has created a pipeline for talented young scientists to engage with these topics beyond the competition (Vinke, Rais, and Millett 2022). When assessing safety and security concerns, iGEM goes beyond traditional, pathogen-based risks, and also considers emerging risks from new technologies. These risks are assessed and monitored throughout the project life cycle, from conception to final presentation (Millett et al. 2019).

3.4.3 Looking ahead

In recent years, there have been renewed calls to address the potential negative impacts of synthetic biology. Thomas Dixon and colleagues have called for a global forum to comprehensively discuss synthetic

biology's societal implications (Dixon et al. 2022); such discussions should include how to de-risk synthetic biology. Effective policy responses need to address the convergence of life sciences, information sciences, and engineering which characterises synthetic biology and synthetic virology (Dixon et al. 2022; National Research Council 2014). Furthermore, artificial intelligence will likely play an increasing role in advancing synthetic biology capabilities.

Advancing capabilities continue to surface new security challenges. This chapter has provided a first overview of evolving capabilities of viral engineering, capabilities with significant security implications. In Part II of this thesis, I deepen this analysis through three case studies of relevant research with dual-use risks.

PART II: CASE STUDY ANALYSIS

Aims and methods

In Part 1, I presented past debates about dual-use research and existing oversight approaches, which primarily have focused on a limited range of microbiological experiments. However, other areas of life sciences and biotechnology research also involve the study and engineering of viruses. In the following chapters, I examine four research areas not captured by existing US oversight, but that may play a crucial role in driving viral engineering capabilities.

In Chapter 4, I examine the dual-use risks of wildlife virus discovery and characterisation. These efforts involve the sampling of wild animals to identify viruses that may eventually jump into humans. This research is not captured by existing United States dual-use oversight, despite featuring the direct study and experimentation of potential pandemic viruses.

In Chapter 5, I consider the dual-use risks of modern vaccine technologies based on engineered viruses. The development of such viral vectors for vaccine delivery may draw on classical virological experiments, including ones subject of previous dual-use discussions, but largely involves synthetic biology knowledge not captured by existing dual-use oversight. One particularly interesting subtopic is the development of transmissible vaccines, which features potentially unique dual-use potential.

In Chapter 6, I study the dual-use risks of viruses engineered to deliver therapeutics. I focus on applications to gene therapy but also consider those pertaining to oncolytic immunotherapy, the targeted killing of cancer cells. These areas have not been traditionally associated with dual-use potential; however, increasing exploration of highly sophisticated viral engineering approaches has led to yet largely unconsidered risks.

In Chapter 7, I examine how artificial intelligence might impact on risks of biological misuse and weaponisation. I focus on two classes of artificial intelligence, large language models (LLMs) and biodesign tools. Artificial intelligence capabilities are currently emerging, so they are not yet captured by existing dual-use policies.

I aim to characterise how different research areas drive viral engineering capabilities, what these capabilities look like, and possible risk mitigation measures. To identify risks and risk mitigation strategies beyond those covered by existing debates and regulations, I base my first principle-based analysis guided by the following questions:

1. Does this research advance capabilities to create or enhance pathogens capable of causing widespread harm?
2. What strategies can mitigate risks while minimising negative impact on legitimate research?
3. Are there any alternative approaches that yield the same benefits but feature less risk?

When I use the term "risk", I refer to the combined product of probability and scale of misuse. The probability of misuse depends on how accessible a given dual-use insight is and whether it creates new capabilities for certain actors.[10] The scale of misuse depends on the nature of the dual-use insight and what dangers it may enable, including how much it raises capabilities to do harm compared to existing methods. In this thesis, I mainly focus on the release of (possibly enhanced) pandemic-capable pathogens, which could result in the largest-scale biological risks.

[10] Probability of misuse also assumes that actors with the intention to misuse a given dual-use insight exist. Given historical precedent of bioterrorism and biological weapons, including by the doomsday cult Aum Shinrikyo, I assume there are actors interested in attaining the capability to start a pandemic.

Chapter 4: Wildlife virus discovery and characterisation[11]

Wildlife virus discovery and characterisation is one proposed measure to mitigate the risks of future pandemics (Gray et al. 2021). Between 2009-2020, the United States Agency for International Development (USAID) funded such research through the PREDICT programs (PREDICT Consortium 2014). Since the COVID-19 pandemic, calls for wildlife virus characterisation have increased. The Global Virome Project calls for a $1.2bn effort to identify 71% of global virome (Carroll et al. 2018), and the new USAID DEEP VZN program has allocated $125m to funding related efforts (USAID 2021). However, the benefits and cost-effectiveness of wildlife virus discovery and characterisation efforts are debated (Holmes, Rambaut, and Andersen 2018). Furthermore, this research involves the direct study and identification of potential pandemic viruses and hence features dual-use risks. These dual-use risks have seen little discussion to date despite being relatively obvious.

I first present the proposed benefits of wildlife virus discovery and characterisation (4.1) and present the relevant research pipeline (4.2), before analysing associated dual-use risks and possible risk mitigation approaches (4.3-5). I argue that the potential identification of pandemic-capable agents features significant security risks and dual-use computational tools are an area of emerging importance for governing risks. I conclude with exploring pathogen surveillance focused on the human-animal interface as a low dual-use alternative to prevent future zoonotic pandemics (4.6).

[11] This section presents part of the content of the following pre-print: Sandbrink, Jonas, Janvi Ahuja, Jacob Swett, Gregory Koblentz, and Claire Standley. 2022. 'Mitigating Biosecurity Challenges of Wildlife Virus Discovery and Characterisation'. SSRN Scholarly Paper ID 4035760. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4035760.

4.1 Benefits of wildlife virus discovery and characterisation

Virus discovery and characterisation efforts aim to identify what viruses could spillover into humans in the future, based on analysing their characteristics and where they are currently circulating (Carroll et al. 2018). Proponents argue that this research features many possible benefits. Identifying high risk viruses may guide clinical assessment and facilitate the detection of pathogen spillover (Carroll et al. 2018). For instance, PREDICT located Marburg virus in Sierra Leone, which now informs local doctors when diagnosing viral hemorrhagic fevers (Amman et al. 2020). Furthermore, proponents argue that identifying future viral threats may facilitate the development of vaccines and therapeutics (Grange et al. 2021). Identified animal viruses may be used to test broad-spectrum countermeasures such as antivirals (Sheahan et al. 2020; 2017; Rappazzo et al. 2021). Additionally, increasing the understanding of viral ecology might inform veterinary medicine or yield other, yet unappreciated, benefits.

Despite these proposed benefits, experts debate whether wildlife virus discovery is cost-effective or even useful for pandemic prevention (Gray et al. 2021; Holmes, Rambaut, and Andersen 2018). Because discovery efforts uncover a vast number of potential animal viruses and a complex set of factors determine their potential to jump species, it is difficult to identify the true danger emanating from discovered viruses (C. J. Carlson 2020). The benefits from PREDICT and other existing studies are difficult to quantify. The $207m PREDICT program identified one new virus that was concluded to have zoonotic potential, which was a result from analysing clinical samples and not from wildlife studies (Steffen et al. 2013). Furthermore, since the SARS outbreak in 2003, animal coronaviruses have been an important part of wildlife virus discovery efforts; arguably, the resulting knowledge has had a limited effect on avoiding the COVID-19 pandemic and informing vaccine design (C. J. Carlson 2020). Instead, crucial information for vaccine design came from MERS-CoV and SARS-CoV after these viruses had caused significant numbers of human infections (Corbett et al. 2020). Lastly, new fast-response vaccine platforms like RNA vaccines

reduce the need for anticipating viral threats, as they enable the development of a vaccine for a newly identified virus within days (Monrad, Sandbrink, and Cherian 2021). Research on prototype pathogens across viral families is important to enable the rapid application of vaccine platforms to new pathogens. Viral discovery efforts are not crucial for advancing this prototype pathogen work, as relevant pathogens across viral families are already known - frequently from previous human infections, as in the case of MERS-CoV and SARS-CoV.

4.2 The wildlife virus discovery and characterisation pipeline

Wildlife virus discovery and characterisation involves a pipeline of diverse research approaches, including fieldwork, computational analysis, and laboratory experiments (Figure 4.1). Researchers collect animal droppings and capture wildlife like bats in their natural habitat to collect faecal, mucosal, or blood samples. The genetic material within these samples is sequenced in laboratories or in the field, and computational methods are used to identify viral genomes. Depending on the project, raw sequencing data or uncharacterised viral genomes may be uploaded straight to public online databases, or the respective viruses might be further characterised. Researchers increasingly use computational tools to identify high-risk animal viruses for further laboratory characterisation, which may involve experiments in cell lines or animal models to assess their potential to infect humans. For instance, Ge *et al.* published the genomes of two SARS-like coronaviruses found in bat faecal samples alongside data on their replication in human cells (Ge et al. 2013). To collate data across viruses and studies, researchers have created an open-source tool that ranks animal viruses by their potential to spillover into human populations (Grange et al. 2021). Findings from viral collection and characterisation inform further wildlife sampling efforts (Figure 4.1a).

Wildlife virus discovery possibilities are expanding due to new and cheaper approaches for sequencing and informatic data interpretation. Metagenomic sequencing allows the non-targeted sequencing of all genetic

material within a given sample. As sequencing costs fall, new viruses may be discovered faster than they can be analysed in the laboratory (Ladner 2021). Thus, new computational tools, including machine learning approaches, are increasingly used to identify the highest-risk viral candidates for further experimental characterisation (C. J. Carlson et al. 2021; Mollentze, Babayan, and Streicker 2021). In turn, data from experimental characterisation studies inform the training of computational models (Figure 4.1b).

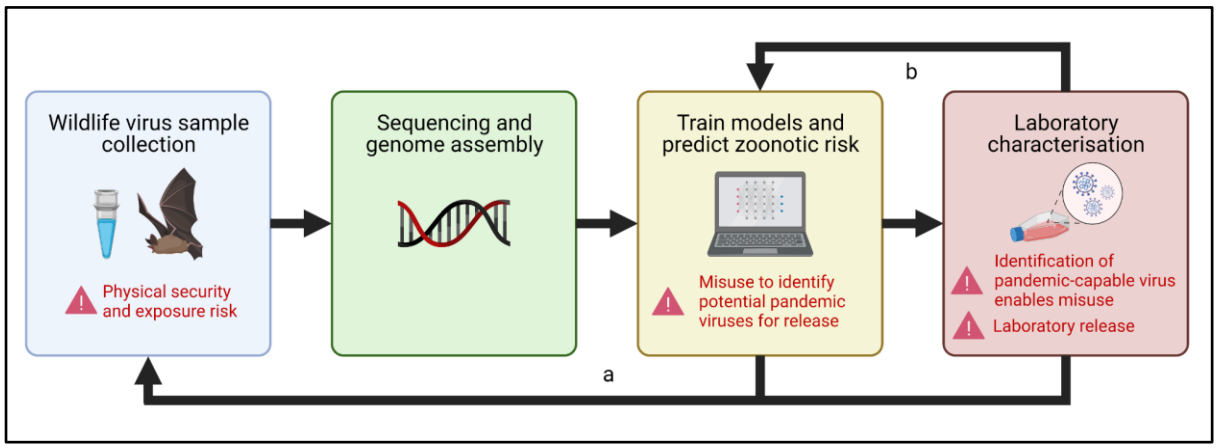


Figure 4.1**:** The wildlife virus discovery and characterisation pipeline and associated risks

Samples collected from wildlife are sequenced and computationally assembled to whole viral genomes. Increasingly, these viruses are computationally characterised and viruses perceived to pose a high risk for infecting humans are studied in laboratories. Findings from experiments then in turn are used to inform further sampling efforts (a) and can help train computational tools for viral characterisation (b). Different biosecurity risks arise throughout this pipeline. Adapted from (Sandbrink, Ahuja, et al. 2022), which in turn adapted this figure from (C. J. Carlson et al. 2021).

Dual-use risks emerge throughout this pipeline (Figure 4.1). I first discuss the dual-use risk of identifying a virus as pandemic-capable in section 4.4, before then discussing dual-use computational tools in section 4.5, and briefly touching on physical dual-use products in section 4.6.

4.3 Dual-use risks from the identification of pandemic-capable pathogens

The identification and dissemination of the blueprints for pandemic-capable viruses features potential for misuse. Despite a similar debate around the products of research to enhance potential pandemic pathogens such as mammalian transmissible highly pathogenic avian influenza virus (Inglesby and Relman 2016), the dual-use potential of efforts to identify pandemic-capable viruses from nature has seen little consideration. The identification of pandemic-capable viruses is particularly concerning as capabilities to synthesise viruses are becoming increasingly widespread and accessible. Kevin Esvelt argues that we are currently partially protected by our lack of knowledge about pandemic-capable viruses against which we do not have any countermeasures (Esvelt 2022). However, as soon as virus discovery efforts publicly identify a pandemic-capable pathogen and publish its genetic blueprint, this potentially creates a great global security risk by handing many individuals and groups the ability to start a pandemic. Thus, if wildlife virus discovery would actually be likely to identify new pandemic-capable pathogens, this research would feature a similar level of misuse risk as ePPP experiments that aim to evaluate how known pathogens might mutate into pandemic-capable agents - as discussed in Chapter 1, these ePPP experiments are subject to oversight in the United States under the P3CO policy. [12]

The global security risk from the public identification of a pandemic-capable virus is not limited to the direct release of such an agent. While extremely concerning, only a small number of potential actors might actually be motivated to start a catastrophic pandemic. However, many more groups and individuals might be interested in advancing their goals by threatening the release of a pandemic-capable agent. To imagine the impact on global security, we might consider the implications of a terrorist group acquiring a nuclear weapon and threatening its detonation. The effects of handing out the credible capability to start

[12] Arguably, while the risks of wildlife virus discovery and ePPP experiments are similar, the benefits of discovering a virus that already exists in nature is in expectation higher than that of creating a putative future pandemic-capable pathogen. Therefore, on the basis of benefits a different treatment could be warranted.

a pandemic would be similarly destabilising, with the difference being that anyone with basic molecular biology knowledge could credibly replicate such threats. The only salvation in such a scenario would be eventual population-wide vaccination against the agent in question. If no pandemic-capable viruses have been publicly identified, a threat would not be credible and much less plausible.

The risk of wildlife virus discovery and characterisation leading to the credible identification of a pandemic-capable virus is tightly linked to its proposed potential to identify such future threats. The magnitude of the dual-use risk of identifying a pandemic-capable virus depends on how likely it is that one animal virus of the 40,000-1,670,000 projected unknown viral species is pandemic-capable (C. J. Carlson et al. 2019; Carroll et al. 2018). Despite the large number of unknown viruses, it seems unlikely that a virus that has never been exposed to humans would feature high human-to-human transmissibility. However, the major proposed benefit of wildlife virus discovery efforts - to "predict" the sources of future pandemics - rests on this assumption. Thus, the inherent dual-use risk of predicting future pandemic viruses highlights a dichotomy: either virus discovery and characterisation efforts to predict the source of future pandemics feature tremendous security risks, or associated benefits for preventing future pandemics are lower than claimed. In either case, this questions whether wildlife virus discovery and characterisation should become a major part of our pandemic prevention and preparedness portfolio. I will discuss alternative pathogen surveillance strategies that may feature both greater benefits for pandemic prevention and less risks in section 4.6.

An actionable strategy to mitigate risks while retaining the majority of benefits might be to refrain from a small subset of virological experiments that drive pandemic-capable virus identification. These experiments include studies of human receptor binding affinity, cell entry, and replication in relevant human cell types, and tests of transmission in appropriate animal models. If an animal virus performs in

these experiments similar to that of a virus of the same family endemic in humans, this would increase the chance that a virus might start a pandemic if introduced into humans. As this small number of experiments can create the strongest evidence that a discovered virus is pandemic-capable, they feature the greatest risks to create dual-use insights. Therefore, refraining from these experiments would remove the bulk of security concerns of identifying a pandemic-capable virus. These experiments are only a small fraction of overall virology, and thus many other potential benefits of wildlife virus discovery and characterisation work would be retained. These would include using identified animal viruses to test broad-spectrum countermeasures, creating phylogenetic trees for rapid characterisation of new threats, and informing local clinical diagnosis and behavioural interventions.

Another strategy for mitigating the misuse of pandemic-capable viruses would be to create systems for the responsible access to relevant pathogen genomes. Currently, all genomic data is shared publicly on platforms like GenBank, European Nucleotide Archive, DNA Data Bank of Japan, and GISAID.[13] However, genomes for potential pandemic pathogens feature significant dual-use risks and thus could be made accessible only to legitimate researchers for approved projects. In an article on the intersection of biosecurity and open science, James Smith and I argue for responsible access to dual-use materials and highlight that similar systems already exist for patient data (Smith and Sandbrink 2022). To access this patient data, researchers need to prove their credentials and need to apply with a concrete project. It seems reasonable to argue that a blueprint for a pandemic agent should be subject to at least the same level of scrutiny. Indeed, individual potential pandemic virus genomes seem like a high leverage point for controlling risks as usually only a small number - certainly less than 100 - laboratories work on a given

[13] These platforms require signing up and agreeing with a tick box to a user agreement that extends to the responsible use of the data.

animal virus. Furthermore, only a part of the viral genome may be required for research, including the development of vaccines or other countermeasures.

4.4 Dual-use potential of computational tools for sequence-function prediction

Computational tools to predict how likely a given discovered virus may spillover and become transmissible in humans are becoming an increasingly important part of wildlife virus characterisation (Ladner 2021). Sequence interpretation has shown increasing success for identifying viruses with the potential to infect humans (Zhang et al. 2019; Kou et al. 2021; Bartoszewicz, Seidel, and Renard 2021). For instance, Mollentze *et al.* have developed a machine-learning tool that interprets signatures of host range in viral genomes to predict their potential for human infection (Mollentze, Babayan, and Streicker 2021). However, learning to predict viral properties based on genomes is associated with dual-use risks.

First, increasingly powerful sequence-to-function predictions may enable malicious actors to identify those viruses that might be pandemic-capable from publicly available genomic data. Once computational tools can identify viruses to be pandemic-capable with even a low level of accuracy and the synthesis of viruses has become even easier, a malicious actor intent on causing a pandemic might simply release multiple computationally identified candidates. Given the uncertainty about the pandemic potential of computationally identified pathogens and greater resources required to create a large set of viruses, misuse risks are likely lower compared to a laboratory-identified virus.

Second, sequence-to-function prediction may eventually enable reversal to enable function-to-sequence prediction. Such pathogen genome optimisation based on desired characteristics would feature significant misuse potential. Malicious actors could leverage such tools to create novel blueprints for viruses optimised for transmissibility and pathogenicity (O'Brien and Nelson 2020). Current models are still some way away

from being reversible. Bartoszewicz *et al.* tried to reverse their sequence-to-function prediction tools and found that artefacts and other constraints prevented function-to-sequence prediction (Bartoszewicz, Seidel, and Renard 2021). Nevertheless, relevant scientific and security communities need to start considering how to mitigate the misuse of future function-to-sequence prediction capabilities, including ones advanced based on computational tools for wildlife virus characterisation.

The growing dual-use risks of computational tools may be mitigated by security-sensitive design and access controls. First, developers could design sequence-to-function prediction tools in a way that makes their unauthorised reversal for function-to-sequence prediction difficult. The current hurdles to the reversal of these tools highlighted by Bartoszewicz *et al.* may serve as a guide (Bartoszewicz, Seidel, and Renard 2021). Second, similar to potential pandemic pathogen genomes, dual-use computational tools that may allow the identification or generation of blueprints for pandemic-capable viruses could be subject to access controls. One prominent tool for managing access to computational tools are application programming interfaces (APIs), the interfaces through which different computer programs interact with each other. For instance, the artificial intelligence research company OpenAI deployed an API to control queries to its powerful language model GPT-3 to mitigate its use for unintended and harmful purposes (Open AI 2020). Access controls to computational tools may learn from DNA synthesis screening. DNA synthesis screening involves both review of customer identity as well as screening of individual synthesis orders. Similarly, API access to computational tools could be restricted both to legitimate customers and sanctioned applications. Major DNA synthesis companies have established the International Gene Synthesis Consortium to coordinate their security measures (Diggans and Leproust 2019). A similar consortium could be formed by the developers of dual-use computational tools or hosts for such tools, like the code-sharing platform GitHub. Computational tools for sequence-to-function predictions are currently developed in a more distributed manner by different academic groups and smaller organisations.

This will complicate establishing a responsible access regime. Nevertheless, as the dual-use risks from these tools will grow, the importance of such access controls will continue to rise.

4.5 Laboratory biosecurity and accident risks

While dual-use knowledge is increasingly becoming the principal security concern of virological research and is the focus of this thesis (Lewis et al. 2019), laboratory biosecurity, preventing the theft or misuse of pathogens from research institutions, remains an important factor for preventing the misuse of dual-use products (Atlas and Dando 2006). Physical biosecurity measures are frequently linked to biosafety interventions to reduce accidental pathogen exposure or sample loss. Wildlife virus discovery and characterisation risks the deliberate or accidental release of viruses during sampling, transport, and laboratory study. Safe sample collection from animals is a technical challenge (Patlovich et al. 2015). Thin gloves allowing the manual dexterity required for animal sampling are easily pierced by the sharp claws of struggling animals. Furthermore, the lack of universal standards for the secure storage and transport of samples risks accidents and theft (Monrad and Katz 2020). The need for transporting dangerous agents to laboratories has been reduced by portable sequencing devices that have enabled the sequencing of samples in the field (Ciuffreda, Rodríguez-Pérez, and Flores 2021). Even in well-controlled laboratories, accidents occur and samples may be mispurposed. In the United States between 2004-2010, a regulated agent was lost in one instance during shipment and high-risk pathogens caused four laboratory acquired infections in BSL-3 facilities (Henkel, Miller, and Weyant 2012).

A large set of possible measures can improve physical biosecurity and reduce accident risks during wildlife virus characterisation and discovery. Foremost among these, universal guidelines for pathogen handling in the field and during transport could both reduce accidental exposures and theft of samples. For instance, the new One Health High Level Expert Panel (OHHLEP) may take the lead on relevant new guidelines

(World Health Organisation 2021a). Such guidelines could draw on practices developed and used for PREDICT's wildlife sample collection efforts (PREDICT One Health Consortium 2016). Laboratories working on relevant potential pandemic agents should follow the new ISO 35001 standard for laboratory risk management (Callihan et al. 2021). Furthermore, researchers aiming to study highly pathogenic or transmissible pathogens should preferentially use safer experimental strategies, like using pseudotyped non-human viruses. Pseudotyped viruses feature the body of a harmless virus in which a protein of the virus in question is inserted, so that its properties and evolution can be studied. This strategy has for instance been used for work on SARS-CoV-2 (Millet et al. 2019). Using pseudotyped viruses more broadly would reduce both physical security and accident risks.

Generally, interventions to reduce the misuse of the physical dual-use products of research are more straightforward and already more broadly adapted than strategies to mitigate the misuse of dual-use knowledge and insights. This broader adoption is reflected in the existence of different biocontainment levels, institutional biosafety boards, and national policies on laboratory risk mitigation and export controls (Casadevall and Relman 2010; Millett and Rutten 2020). While an important topic, further discussion of mitigating the misuse of physical research products is out-of-scope for this thesis. I focus on the challenge of managing dual-use insights and only discuss physical research products to a limited extent.

4.6 Low-risk alternatives to wildlife virus discovery and characterisation

Critics of wildlife virus discovery and characterisation have previously highlighted the limited benefits of this work for pandemic prevention. Virus populations are constantly changing and the factors determining spillover into humans are very complex, thus assessing future threats based on wildlife sampling is incomplete and biased (Wille, Geoghegan, and Holmes 2021). Furthermore, even if researchers identify a virus that might at some point spillover into humans and cause a pandemic, the potential for

intervention is limited. COVID-19 has demonstrated that it is difficult to vaccinate populations against a virus already spreading through the human population (Razai et al. 2021); it would be substantially more difficult to vaccinate against possible future threats. As wildlife virus discovery and characterisation features significant risks and has questionable benefits, the research community should actively consider low-risk alternatives to such work.

One such alternative strategy might be increased investment into detecting and investigating spillover events at the human-animal interface. The risk-benefit calculation of dual-use insights on a pandemic-capable pathogen hinges on whether a pathogen has already been introduced into humans or not. If a pathogen is circulating in humans, then the impact of an additional deliberate release would be substantially reduced, while the public health importance of sharing this information would be significantly greater. Therefore, the detection of novel pathogens in humans may be both lower risk and more cost-effective for preventing pandemics than attempts to predict future threats (Holmes, Rambaut, and Andersen 2018).

Zoonotic pathogens usually require multiple spillovers into humans before acquiring human-to-human transmissibility (Gray et al. 2021). Thus, identifying viruses that regularly make the jump into humans before they acquire human-to-human transmissibility may be much more important for pandemic prevention than screening for possible future threats in animal populations (Geoghegan and Holmes 2017). Surveillance of high risk occupations and locations for infection with novel pathogens is thus important for pandemic prevention (Wille, Geoghegan, and Holmes 2021). To this end, Africa CDC has started the Sentinel program which uses a combination of pathogen-specific and pathogen-agnostic testing for providing early warning of possible pandemic events (Africa CDC 2020).

Once a zoonotic spillover has been documented, targeted animal sampling may be used for follow-up. Such investigations may for instance uncover the animal reservoirs of the pathogen in question, to then inform risk communication and behavioural interventions (Saylors et al. 2021). Focusing animal sampling on the human animal interface would thus ensure that any findings are maximally actionable - and would minimise the risk of uncovering pandemic-capable pathogens that might never make it into human populations.

4.7 Lessons from assessing wildlife virus discovery and characterisation

Considering the dual-use risks of wildlife virus discovery and characterisation highlights multiple lessons. First, the identification of pandemic-capable viruses from nature features grave global security implications. Associated dual-use risks may be of similar scale to the dual-use risks of ePPP research, as both lines of research may identify new agents potentially capable of causing pandemics. While ePPP research has recognised dual-use risks and is subject to oversight in the US, relevant wildlife virus discovery research is not widely considered to feature substantial dual-use risks. Second, benefits and risks of the ambition to predict future pandemic viruses - whether actually possible or not - are tightly coupled. This research is conducted to understand the threats that might emerge from nature, however the same knowledge enables deliberate release. To reduce the risk of identifying a pandemic-capable pathogen, researchers would have to refrain from conducting a select handful of viral characterisation experiments. This would retain the bulk of other benefits of wildlife virus discovery and characterisation. Instead of attempting to predict future viral threats, alternative pandemic prevention investments feature less risks and potentially greater cost-effectiveness - one example being improved testing of high-risk individuals for infections with novel pathogens. Lastly, computational tools will likely play an increasing role in driving dual-use capabilities. Responsible access systems could be useful for controlling access to dual-use genetic blueprints and computational tools.

Chapter 5: Virally vectored vaccines[14]

Vaccines are one of the most established applications of engineered viruses. Historically, vaccine facilities for the large-scale culture of bacteria and viruses have been associated with dual-use potential. However, classical inactivated and attenuated pathogen vaccine approaches are increasingly being displaced by new approaches to create vaccines that rely on advances in synthetic biology. One new vaccine platform are viral vectors, non-pathogenic viruses that are engineered to induce immunity against a desired pathogen. The AstraZeneca and Johnson & Johnson COVID-19 vaccines have been based on this technology (Falsey et al. 2021; Sadoff et al. 2021). Thus, in the wake of the COVID-19 pandemic, there is potential for increased investments into exploring synthetic viruses as vaccine vectors (Monrad, Sandbrink, and Cherian 2021). I chose this field as a second case study, because engineering viral vectors as vaccines may inadvertently advance capabilities to engineer viral pathogens but these risks have not previously been properly characterised.

First, I provide background on virally vectored vaccines and their benefits and downsides as vaccine platforms (5.1). Then I present how dual-use considerations around vaccine development have shifted from the misuse of equipment to intangible insights (5.2) and evaluate which insights feature the greatest dual-use potential (5.3). In a separate section, I introduce work on transmissible vaccines and their unique ethical challenges as self-spreading agents associated with unique dual-use insights (5.4). I highlight how the preferential advancement of low-risk avenues can reduce dual-use risks from virally vectored vaccines (5.5).

[14] This section draws heavily on the content of two papers I wrote before the DPhil:

1. Sandbrink, Jonas B., and Gregory D. Koblentz. 2022. 'Biosecurity Risks Associated with Vaccine Platform Technologies'. *Vaccine* 40 (17): 2514–23. https://doi.org/10.1016/j.vaccine.2021.02.023.
2. Sandbrink, Jonas B., Matthew C. Watson, Andrew M. Hebbeler, and Kevin M. Esvelt. 2021. 'Safety and Security Concerns Regarding Transmissible Vaccines'. *Nature Ecology & Evolution* 5 (4): 405–6. https://doi.org/10.1038/s41559-021-01394-3.

5.1 Strategies and properties of virally vectored vaccines

Virally vectored vaccines are created by inserting a genetic element of the vaccine target into a well-defined, non-pathogenic virus. For instance, the ChAdOx1 Oxford/Astrazeneca COVID-19 vaccine consists of a SARS-CoV-2 spike gene inserted into a chimpanzee adenovirus vector (Figure 5.1a) (Folegatti et al. 2020). The resulting engineered viral vector is then injected into the patient, enters cells, produces the spike protein, and induces an immune response against this vaccine target (Rauch et al. 2018).

A large range of viral vectors have been explored as vaccine vehicles and feature different upsides and downsides (Ura, Okuda, and Shimada 2014). Adenoviruses are one of the most popular vectors, having not just featured in ChAdOx1 but also the Johnson & Johnson COVID-19 vaccines (Sadoff et al. 2021). Adeno-associated virus (AAV) is a popular viral vector and while it is mainly used for gene therapy due to the low level of inflammation it induces, it has also been explored for vaccines (Xin et al. 2006; Lin et al. 2009). Notable other viruses explored as viral vectors for vaccines are vesicular stomatitis virus (VSV), used in an Ebola vaccine (Henao-Restrepo et al. 2017), and vaccinia virus, one of the oldest viral vectors which has been explored for a range of applications (Kaplan 1989; Rerks-Ngarm et al. 2009).

Virally vectored vaccines are an example of a platform vaccine technology. Platform vaccines are systems where the same well-characterised delivery system can be easily adapted to target a new pathogen (Adalja et al. 2019). Next to viral vectors serving as such vaccine platforms, RNA vaccines, DNA vaccines, and a subset of protein expression systems may be considered platform approaches. Because of their potential to be rapidly adapted to a new target, the advancement of platform vaccine technologies may be a promising strategy to prepare for future pandemics (Monrad, Sandbrink, and Cherian 2021). For instance, the ChAdOx1 virally vectored vaccine for COVID-19 was based on a vector previously studied for a MERS

vaccine, for which safety studies already had been conducted; this accelerated large-scale clinical trials (Barker 2020; Bosaeed et al. 2022). Among platform vaccine technologies, virally vectored vaccines are among the more well-defined technologies. Even before COVID-19, virally vectored vaccines had been licensed for Ebola and Dengue (Sebastian and Lambe 2018).

Next to their potential for rapid development, virally vectored vaccines feature other unique characteristics. They may be particularly effective at inducing not only humoral but also strong cellular immune responses, thus featuring particular benefits for pathogens with complex pathogenesis (Graham and Sullivan 2018). Furthermore, as the Johnson & Johnson COVID-19 vaccine demonstrated, virally vectored vaccines may be used in a single dose regimen (Sadoff et al. 2021). Additionally, compared to RNA vaccines, virally vectored vaccines have lower cold chain requirements (Falsey et al. 2021; Polack et al. 2020). Lastly, virally vectored vaccines may be engineered as transmissible vaccines, which are being explored for the vaccination of animal populations. I discuss these transmissible vaccines and their dual-use potential separately in section 5.5.

Virally vectored vaccines are limited by anti-vector immunity. Pre-existing immunity against the vaccine vector, induced by previous natural infection with a related virus, may decrease the effectiveness of virally vectored vaccines (Saxena et al. 2013). Antibodies may neutralise the viral vectors before they can enter cells and express the vaccine target. Adenovirus 5 (Ad5) is a viral vector with promising properties but due to its natural existence as a common cold virus, immunity against Ad5 is widespread (Pichla-Gollon et al. 2009; Pine et al. 2011). Furthermore, reusability of a given viral vector may also be limited as any administration may not only induce immunity against the vaccine target but also the viral vector (Saxena et al. 2013). Accordingly, different approaches are taken to evade anti-vector immunity (Figure 5.1); I will discuss these approaches and their dual-use risks in more detail in 5.4 and 5.5.

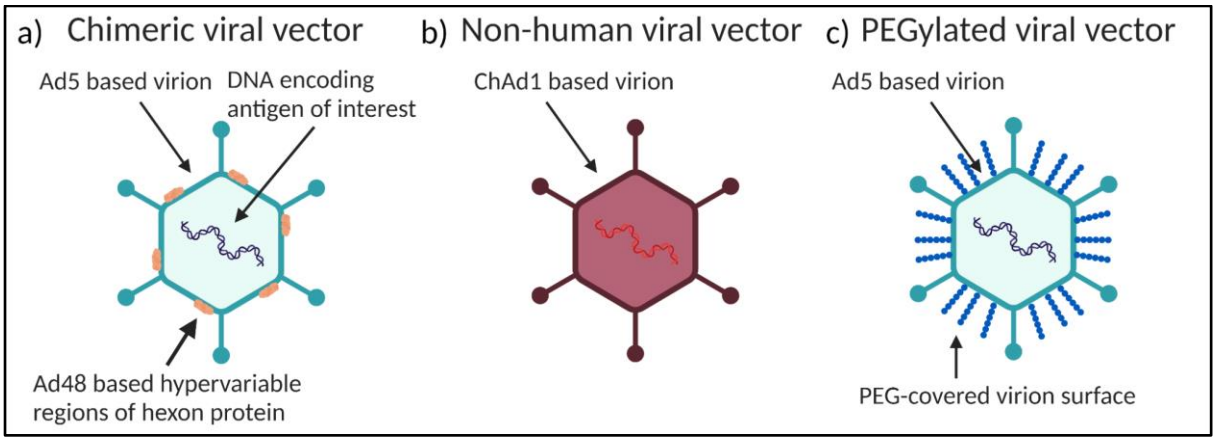


Figure 5.1: Viral vectors and their modifications for vaccines

Virally vectored vaccines consist of a viral vector with a genome in which the gene for the antigen of interest, a protein of the vaccine target, has been inserted. Different approaches to overcome pre-existing anti-vector immunity are a) the creation of chimeric viral vectors with surface proteins that do not feature pre-existing immunity, b) choosing virus as vectors that do not feature pre-existing immunity, c) chemical modification of viral surfaces, for instance with polyethylene glycol (PEG). Figure 5.1 was created using biorender.com. Reproduced from Sandbrink and Koblentz 2022.

5.2 From physical security risks to dual-use insights

The dual-use potential of vaccines has evolved over time with technological advances. Traditionally, manufacturing equipment for vaccine production was associated with significant dual-use potential. Fermenters for the large-scale culture of viruses used in attenuated or inactivated vaccines could be misused for the large-scale production of agents for biological warfare. For instance, Iraq used fermenters originally intended for producing veterinary vaccines for the production of botulinum toxin (Atlas and Dando 2006; Central Intelligence Agency 2005). Similarly, the Soviet Union planned to repurpose vaccine production plants for production of biological warfare agents (Leitenberg 2001).

Increasingly, dual-use insights derived from research on viruses are displacing manufacturing equipment as the principal source of security concern. As mentioned in Chapter 1, a particular newsworthy case

demonstrating this development was the debate ensuing after the recreation of horsepox virus as a new vaccine vector (Noyce, Lederman, and Evans 2018; Kupferschmidt 2018). Biosecurity experts argue that this public demonstration of how to stitch together a poxvirus from synthetic DNA has lowered the barrier to acquisition of smallpox, otherwise only held at two secure repositories in the United States and Russia (Koblentz 2017; Inglesby 2018).

Advances in synthetic biology, including cheaper DNA synthesis, increase the threat from the misuse of dual-use insights. While large-scale production of classical biowarfare agents, including botulinum toxin and anthrax, would be associated with large but limited harm, the release of a modified, highly transmissible virus might pose a global catastrophic risk. As synthetic biology is becoming more accessible and powerful, the creation and modification of such agents may be possible. Public information, including information arising from viral vector research, once published may be misused by a large number of individuals around the globe. Thus, dual-use insights from research on virally vectored vaccines may feature high dual-use potential compared to other vaccine technologies (Sandbrink and Koblentz 2022).

5.3 Dual-use viral engineering approaches

There are two main categories of dual-use risks arising from work on virally vectored vaccines: a) the general spread of skills and methods to create viruses and b) specific highly dual-use insights on viral engineering.

Any work involving the synthetic creation of viruses will advance relevant dual-use methods and make these accessible to a greater number of individuals. Creating virally vectored vaccines advances molecular biology methods to create and amplify synthetic DNA and trigger viral assembly in cell lines. While featuring dual-use potential, these general-purpose skills are driven by a large range of research. Based on

bibliometric data, I estimate that only 10% of all research on viral vectors is vaccine research (Sandbrink and Koblentz 2022). A large fraction of other research on viral vectors is conducted for gene therapy. However, vaccine work may drive a particularly dangerous subset of synthetic biology skills. Viral vectors used in vaccines are frequently related to pathogens and thus may inform misuse. For instance, as the horsepox debate demonstrated, work on poxviruses like vaccinia virus and horsepox as vaccine vectors may uniquely spread skills for the recreation of smallpox.

The greatest dual-use risks are however associated with developing specific approaches for enhancing properties critical for pathogens, most notably immune evasion. Anti-vector immunity poses a challenge for virally vectored vaccines: this incentivises a large set of research on engineering immune evasion (Thacker, Timares, and Matthews 2009). To realise the promise of Ad5 as a vaccine vector despite widespread pre-existing immunity, Roberts *et al.* have developed a chimeric adenovirus candidate with surface proteins from Ad48, an adenovirus serotype with little pre-existing immunity (Roberts et al. 2006) (Figure 5.1a). Such approaches are particularly worrying when applied to vectors related to pathogens with pandemic potential, including influenza virus, poxviruses, measles virus, and poliovirus. These methods may be misused to turn these pathogens into versions capable of circumventing natural or vaccine-acquired immunity and thus able to spark pandemics. Surface protein engineering is generally virus family specific which limits the misuse potential of work on vectors unrelated to concerning pathogens. However, in the future, academics and companies may explore more sophisticated immune evasion approaches with greater universality and, thus, greater dual-use potential.

5.4 Dual-use risks of transmissible vaccines

Research on transmissible vaccines is associated with particular ethical challenges, including considerations around dual-use risks. Researchers have proposed to use such transmissible vaccines for

animal populations to target pathogens that might spillover into humans. The most popular version of transmissible vaccines are virally vectored vaccines where transmissibility of the viral vector is retained, thus this viral vector would transmit from the first exposed animals and induce immune responses against the target virus in every encountered animal. Transmissible vaccines have the upside of enabling the potential cost-effective immunisation of large populations of wild animals. Thus, they could have a direct application for diseases that frequently enter humans from known reservoirs, like rabies and Lassa fever (Griffiths et al. 2022; Nuismer 2022). Additionally, proponents argue that transmissible vaccines could eventually preemptively eradicate high risk zoonotic pathogens that feature pandemic potential (Nuismer and Bull 2020). While one field trial of a transmissible vaccine for rabbit myxomatosis and rabbit hemorrhagic disease was conducted in the early 2000s, the development of viral vector-based transmissible vaccines is mainly in an early stage (Griffiths et al. 2022). However, funding bodies, including the United States Defense Advanced Research Projects Agency (DARPA), are interested in advancing this technology.[15]

The development and deployment of transmissible vaccines features many ethical challenges beyond dual-use risks. Once released, transmissible viruses may evolve and inadvertently cause harm (Lentzos et al. 2022)**.** Even viruses with a low reproductive rate may undergo genetic changes. This is demonstrated by the live (Sabin) poliovirus vaccine, which transmits only infrequently but nevertheless regularly undergoes genetic changes and reverts to a pathogenic form (Burns et al. 2014). While a loss of the vaccine genes is the most likely genetic change, recombination events with other viruses may lead to inadvertent changes, including a transfer of genetic material with human viruses (Sandbrink et al. 2021). Additionally, it can be difficult to run controlled trials because of the risk of escape. For instance, a rabbit hemorrhagic disease virus escaped during a presumably secure trial on an isolated island and subsequently killed rabbits on

[15] Based on a conversation with a DARPA contractor.

mainland Australia, before being deliberately smuggled to New Zealand (Schwensow et al. 2014). Some of the ethical challenges of transmissible vaccines have been discussed as part of broader discussions on self-spreading viruses for wildlife management, including liability and responsibility if such agents cross national borders (Lentzos et al. 2022).

Transmissible vaccine development may be associated with particular dual-use insights. While other virally vectored vaccine work may involve the engineering of immune evasion, transmissible vaccines may uniquely generate insights into the fine-tuning of viral transmissibility and genetic stability (Sandbrink et al. 2021). Despite current viral vector research not focusing on this aspect, it seems likely that increased funding for this work would generate an incentive to investigate this topic. Furthermore, research into enhancing the genetic stability of recombinant viruses may be explored to prevent the loss of important, but evolutionarily disadvantageous, genetic vaccine elements. Methods to confer genetic stability may be misused to lock in certain dangerous properties, like enhanced lethality, at a level that would otherwise be lost on sustained transmission. Lastly, any modification of transmissible vaccines has to be able to be passed onto vaccine progeny, reducing the potential for low dual-use non-heritable modification strategies.

5.5 Risk mitigation through low-risk alternatives

Dual-use risks from virally vectored vaccine development can be mainly mitigated through pursuing low-risk alternatives. In contrast to wildlife virus discovery, which, as discussed in Chapter 4, creates many dual-use insights through work directly on pandemic-capable pathogens, the risks of developing virally vectored vaccines are more indirect. Thus, fewer strategies are available to modify risks, such as improving experimental practices and the careful handling of a small subset of information. Instead, the most promising way to mitigate dual-use risks may be to shift away from lines of research that predominantly generate these risks and choose low-risk paths that offer the same benefits.

When developing a virally vectored vaccine, risks can be reduced through choosing certain research avenues over others. Generally, picking viral vectors that are not related to potential pandemic pathogens will reduce dual-use risks. For instance, the development of an adenovirus-based vaccine will likely generate skills and insights creating less misuse risks than work on poxviruses, such as vaccinia virus. To reduce the need of engineering immune evasion, vaccine developers might choose viral vectors without preexisting immunity, such as the chimpanzee adenovirus used in the ChAdOx1 Astrazeneca/Oxford vaccine (Figure 5.1b) (Falsey et al. 2021). The ChAdOx1 and rVSV-based Ebola vaccines demonstrate that non-human viral vectors can be safe and efficacious (Henao-Restrepo et al. 2017; Falsey et al. 2021). Vaccine developers exploring a promising vector stunted by pre-existing immunity might use non-heritable methods over heritable methods for achieving immune evasion. For instance, chemical modifications may be used to confer immune evasion but feature little misuse potential as they are not suitable to enhance transmissible viruses (Figure 5.1c). For example, Weaver and Barry showed that a PEGylated Ad5 vaccine vector induced higher levels of antibodies compared to unmodified Ad5 (Weaver and Barry 2008).

Risks can be reduced the most by moving away from vaccines based on engineered viruses to alternative technologies. RNA vaccines, vaccines based on the genetic messenger RNA delivered in lipid nanoparticles, do not involve any viral engineering and hence feature lower dual-use potential than virally vectored vaccines (Sandbrink and Koblentz 2022). Fortuitously, RNA vaccines showed greater efficacy against COVID-19 than virally vectored vaccines (Polack et al. 2020; Sadoff et al. 2021; Falsey et al. 2021). Before COVID-19, RNA vaccines were still a quite immature technology with questionable promise, however now increased investment and wide application seems likely (Sandbrink and Shattock 2020). Preferentially developing and using RNA vaccines would reduce the build-up of population immunity against the most promising viral vectors and thus might retain the efficacy of virally vectored vaccines for when their unique properties at inducing cellular immune responses are needed, including for poxviruses,

filoviruses, and other enveloped viruses with complex pathogenicity (Graham and Sullivan 2018). A low dual-use alternative to transmissible vaccines based on self-spreading viruses are transferable vaccines (Sandbrink et al. 2021). These are topically applied to animals for spread through close contact and have shown conceptual promise (Bakker et al. 2019). However, the bulk of dual-use risk unique to developing transmissible vaccines may also be removed by a commitment to not enhance vector transmissibility.

5.6 Lessons from assessing dual-use risks of virally vectored vaccines

Analysis of the dual-use potential of developing virally vectored vaccines highlights multiple lessons. First, research on viral vectors generates dual-use viral engineering skills and methods. Much of the most sophisticated viral vector engineering is conducted for gene and cancer therapy, which I examine in the next chapter. Second, the enhancement of viral properties like immune evasion and transmissibility drives a significant fraction of dual-use risks of developing virally vectored vaccines. This highlights the potential for select interventions and advocacy with relevant researchers to shift towards less risky approaches. Lastly, generally, choosing low-risk alternatives are a crucial strategy for reducing risks from developing virally vectored vaccines. Creating systems in which low-risk alternatives are preferentially pursued may be a generalisable risk mitigation strategy for dual-use biotechnology research that does not involve direct work on pathogens.

Chapter 6: Viral vectors for gene therapy[16]

A large amount of research on viral vectors, and thus much of synthetic virology, is driven by gene therapy (Sandbrink and Koblentz 2022). The enduring hope to end genetic diseases sparks commercial interest, which in turn incentivises relevant academic research. This trend is reinforced by the successful licensing of several gene therapies (Shahryari et al. 2019). Gene therapy delivery requires viral vectors with particular properties, some divergent from the original functions of viruses; thus, gene therapy has inspired sophisticated approaches to engineer viral properties. Furthermore, researchers develop general-purpose viral engineering tools to optimise viruses across many properties at once and make these accessible to individuals without extensive virological expertise (Ogden et al. 2019; Sinai et al. 2021). Thus, research on viral vectors for gene therapy delivery may be an important source of emerging dual-use risks. The main focus of this case study are viral vectors for gene therapy. Still, I also touch on related efforts to develop cancer-killing oncolytic viruses and other cancer immunotherapies based on viral vectors.

First, I discuss why and how viral vectors are engineered for gene therapy and in what ways this work differs from previously discussed virally vectored vaccines. Then, I analyse three concrete subsets of dual-use risks (6.2-4), highlighting both how virus-specific work for gene therapy research may be less concerning than that for vaccines and how gene therapy research advances highly dual-use universal methods for viral enhancement. I end with a discussion of how to mitigate these risks, finding that previously discussed strategies of low dual-use alternatives (6.5) and structured access to computational tools (6.6) are also applicable for this research.

[16] This section presents the content of the following paper: Sandbrink, Jonas B., Ethan C. Alley, Matthew C. Watson, Gregory D. Koblentz, and Kevin M. Esvelt. 2022. 'Insidious Insights: Implications of Viral Vector Engineering for Pathogen Enhancement'. *Gene Therapy*, March, 1–4. https://doi.org/10.1038/s41434-021-00312-3.

6.1 Viral vectors and engineering challenges

Viral vectors are currently the most popular method to deliver gene therapies. Similar to virally vectored vaccines, the gene of interest, for example, a gene missing as part of a genetic disorder, is inserted into a non-pathogenic viral vector. Viral vectors are selected for different disorders based on whether they target relevant tissues. For instance, adeno-associated virus serotype 2 (AAV2) can enter cells of the eye and thus has been used in a gene therapy product to treat the eye disease Leber congenital amaurosis (Casey, Papp, and MacDonald 2020). In contrast to vaccines, viral vectors for gene therapy work best when inducing minimal inflammation and immune activation (Chan et al. 2021). Another crucial property of gene therapy vectors is long-term gene expression (Haberman, McCown, and Samulski 1998). Thus, gene therapy frequently features different vectors to those used in vaccines and cancer immunotherapy. Adeno-associated virus has become the most popular viral vector for gene therapy due to inducing little inflammation, providing long-term gene expression, and offering many serotypes with different tissue specificities.

Engineering and recombination of viral vectors can help to optimise their properties and thus improve the effectiveness of gene therapies and suitability for more diseases. Like virally vectored vaccines, viral vectors for gene therapy are limited by anti-vector immunity. Surface protein engineering approaches are employed to achieve evasion of pre-existing immune responses similar to virally vectored vaccines (further discussed in 6.2). Additionally, researchers are also engineering vectors to modulate the immune system to temper cellular inflammation induced by viral genetic material (further discussed in 6.3). Limited potential for readministration due to anti-vector immunity is another related challenge. To optimise the targeting of desired tissues, researchers reconstruct ancestral viral particles (Santiago-Ortiz et al. 2015) and recombine existing vectors (Castle et al. 2016). Lastly, viral vectors are optimised across many other

properties like expression efficiency, ease of production, and thermostability with increasingly powerful and generalisable viral engineering approaches (further discussed in 6.4) (Ogden et al. 2019).

6.2 Dual-use risks from virus-specific enhancement

In my assessment of dual-use risks of viral vectors, I focus on viral engineering insights that may inform the enhancement or creation of pandemic-capable viruses. Others have previously discussed how viral vectors may be misused to deliver toxins or other harmful genes (Kirkpatrick et al. 2018; Javitt and Prince 2012), however these risks are out of the scope of this work due to the non-replicating nature of such constructs and less relevance to informing the creation of globally catastrophic pandemics. Instead, I will focus on those insights which are capable of turning viruses pandemic-capable. Many pathogens are constrained by pre-existing immunity or insufficient human receptor adaptation. Thus, methods for immune evasion or enhancing receptor binding originating from gene therapy research may be used to create new pandemic agents. Particularly concerning are insights directly applicable to viruses capable of autonomous spread.

Pre-existing anti-vector immunity limits promising gene therapy vectors like AAV (Calcedo et al. 2009). Thus, surface proteins of viral vectors for gene therapy are modified using similar approaches as those for vaccines. Researchers study the effects of different mutations of AAV capsids on immune evasion (Lochrie et al. 2006) and immune evasive variants are selected through directed evolution, a process in which large numbers of mutations are generated and screened for the ability to evade neutralising antibodies (Maersch et al. 2010). Resulting AAV vectors may be able to evade pre-existing immunity whilst retaining favourable properties for gene delivery.

Such virus-specific approaches for optimising AAV feature limited dual-use risks. This is because AAV requires co-infection with a helper virus, adenovirus, to replicate in human cells as it does not feature its own replication machinery. Therefore, it is much more difficult to turn AAV into a pandemic pathogen and insights specific to AAV are less concerning (Figure 6.1d) (Meier, Fraefel, and Seyffert 2020). In contrast, similar work on viruses capable of unassisted human transmission features comparatively higher risks (Figure 6.1e). As for vaccines, the risks depend on how easily a given virus might be misused and turned into a pandemic pathogen. Most concerning may be virus-specific cell surface engineering of influenza virus, measles virus, and poliovirus, which are exploited by oncolytic virus research for their cell killing and immune induction abilities (Denniston et al. 2016; Msaouel et al. 2012; Pizzuto et al. 2016).

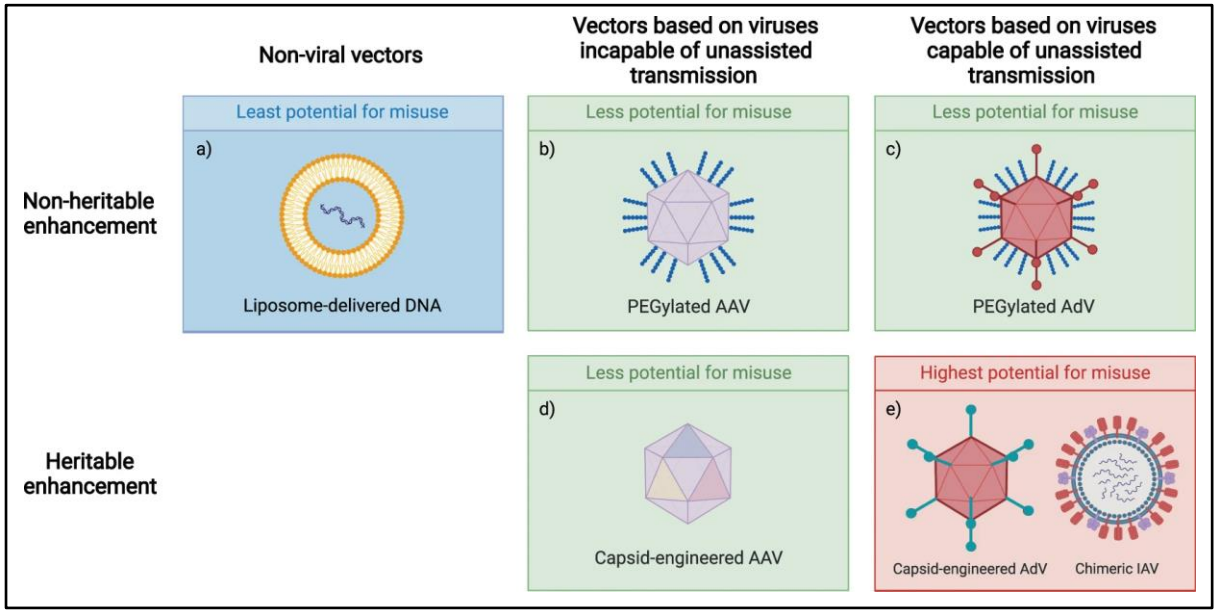


Figure 6.1: Vector enhancement approaches and their relative misuse potential

a) Non-viral vectors do not involve any viral engineering and hence feature the least potential for misuse. b-c) Non-heritable enhancement of viral vectors features comparatively less potential for misuse compared to heritable enhancement. d) Heritable enhancement of AAV, a virus incapable of unassisted transmission, features less potential for misuse than that of viruses capable of unassisted transmission. e) Heritable enhancement of viruses capable of unassisted transmission generally features the highest potential for misuse. Reproduced from Sandbrink et al. 2022.

6.3 Dual-use risks from universal approaches for immune modulation

To inhibit immune activation, researchers have been exploring ways to modify viral vectors with immune-dampening genetic elements. These genetic elements may for instance be taken from other viruses, such as US6 glycoprotein from cytomegalovirus (Kwangseog Ahn et al. 1997) and infected cell protein 47 from herpes simplex virus (K. Ahn et al. 1996). Both genes prevent immune cells from killing infected cells, and have been used to enhance AAV (Shao et al. 2018). However, novel approaches to interfere with immune system activation are also being explored. An example is the study by Chan *et al.,* which I described in the introduction to this thesis. As discussed earlier, Chan *et al.* used short non-coding DNA sequences to interfere with the activation of the TLR9 innate immune sensor to prevent the detection of AAV genetic material and subsequent induction of an inflammatory response (Chan et al. 2021).

The development and characterisation of such genetic elements for immune modulation features high dual-use potential. Their functionality resembles that of work that sparked a well-known dual-use controversy discussed in Chapter 1: Jackson et al. inserted in the early 2000s the immunosuppressive cytokine IL-4 into the genome of mousepox virus, which conferred enhanced virulence and vaccine evasion in mice (R. J. Jackson et al. 2001). Biosecurity experts noted the grave misuse potential of this method due to the similarity between mousepox and smallpox virus (Selgelid and Weir 2010). Similarly, immune modulatory genetic elements for viral vectors for gene therapy may be misused and inserted into pathogens. Indeed, these genetic elements can be inserted into any virus with sufficient genomic flexibility as they operate independently of other viral systems. Concerningly, the short non-coding DNA sequences used by Chan *et al.* are much shorter than the genes for IL-4 or other proteins. Thus, these universal genetic elements may be transferable to viruses with smaller genomes, including many high-risk pathogens. This application may require little specialist expertise or additional research. Current genetic elements for

immune modulation still feature limited functionality, but future advances of this kind may generate very generalisable capabilities to enhance viruses.

6.4 Dual-use risks from general-purpose viral engineering methods

The need to optimise viral vectors for gene therapy across many different properties advances new viral engineering methods. These methods are frequently general-purpose and can facilitate faster or more precise viral engineering experiments, with some even reducing the need for certain experiments. Thus, these general-purpose viral engineering methods may lower the barrier to creating a pandemic-capable pathogen.

Certain general-purpose viral engineering methods advanced through gene therapy research still require work-intensive experiments and specialist expertise, and thus feature a higher barrier to misuse. This includes cutting-edge high throughput methods for directed evolution across multiple properties (Ogden et al. 2019). Structure-guided rational design can also inform vector-enhancing mutations but requires expert judgement and structural data from cryo-electron microscopy or X-ray crystallography experiments (Ding and Gradinaru 2020).

Computational tools for general-purpose viral engineering may feature greater misuse potential. Ogden *et al.* pioneered the computational optimisation of AAV across properties like virus production, immune evasion, thermostability, and biodistribution (Ogden et al. 2019). Other groups have since built on this approach (Marques et al. 2021; Mikos, Chen, and Suh 2021). Computational tools are now able to predict viable capsids that are substantially different from experimentally derived variants (Marques et al. 2021; Bryant et al. 2021). The need for specific experimental training data for machine learning tools has been

removed through demonstrations of how publicly available data can be used to inform the identification of new synthetic variants (Mikos, Chen, and Suh 2021; Sinai et al. 2021).

Computational tools lower the need for expert judgement and cut down on laborious experiments. I will discuss general developments relating to large language models and biodesign tools in Chapter 7. However, there are also tools and applications specific to viral vector design. Papers describing computational advances highlight how these approaches are transferable to other viral engineering challenges (Ogden et al. 2019) and are aiming at enabling a broad set of wet-lab scientists without specific specialist expertise in the enhancement of particular viral properties (Sinai et al. 2021). The development of protein folding prediction tools like AlphaFold may exacerbate these developments, including through allowing the generation of large synthetic training datasets for the most concerning pathogens (Baek et al. 2021; Jumper et al. 2021). Currently, computational tools are still limited in power and transferability between viruses and target properties, but this will change as researchers seek to build increasingly powerful and general-purpose tools.

6.5 Mitigating risks through low-risk alternatives

Similar to virally vectored vaccines, risks from viral vectors for gene therapy can principally be cut through the preferential advancement of low dual-use solutions. There may be other risk mitigation approaches for computational tools, which I will discuss separately in the next section (6.6).

Particularly promising for reducing dual-use risks are alternatives to the heritable enhancement of viral vectors. Similar to vaccines (discussed in 5.5), risks may be reduced by choosing vectors that do not require extensive viral engineering associated with dual-use risks, such as vectors without pre-existing immunity (Jeune et al. 2013). Furthermore, sticking to AAV and virus-specific methods for the genetic modification

of AAV would keep dual-use potential small. For other vectors, non-genetic modifications like shielding through synthetic polymers like PEG (Yao et al. 2017) or the use of lipid bilayer envelopes (Singh, Tian, and Kostarelos 2008) may be a low dual-use alternative to heritable enhancement methods (Figure 6.1b,c).

Other strategies are also available to minimise dual-use risks of viral vector engineering for gene therapy. Instead of designing ever more potent dual-use genetic immune-modulating elements, clinicians could focus on pharmacological means to induce immune modulation. For instance, drugs that prevent the activation of immune responses or induce immune tolerance have been explored as complements to gene therapies (Adriouch et al. 2011; Meliani et al. 2017; Mitchell and Samulski 2013). Novel advances in synthetic biology may also be leveraged to develop genetic enhancement methods that are difficult to misuse. For instance, one might use non-canonical amino acids (de la Torre and Chin 2021), amino acids normally not occurring in the human body, in vector engineering and thus prevent misapplication of such engineering approaches outside of controlled environments where such unnatural amino acids are provided.

In the same way that RNA vaccines present a low dual-use alternative to virally vectored vaccines, non-viral delivery methods might be a potent way to reduce the dual-use potential of gene therapy development (Figure 6.1a) (Sandbrink, Alley, et al. 2022; Sandbrink and Koblentz 2022). For instance, lipid nanoparticles (LNPs) have been explored as delivery vehicles for gene therapies (Cullis and Hope 2017). Non-viral gene therapy methods used to feature poor delivery efficiency but are increasingly the subject of human trials (Ramamoorth and Narvekar 2015). The timelines of replacing viral delivery will likely be longer than RNA vaccines gaining greater popularity than virally vectored vaccines. However, non-viral delivery methods feature advantages like cell-free manufacturing, larger payloads, and flexible design which make them attractive (Cullis and Hope 2017). Importantly, they do not feature the challenge of

anti-vector immunity and limited potential for repeat administrations (Coelho et al. 2013). To achieve non-viral gene therapy delivery, researchers will need to solve the problem of how to target specific cell types; advances in lipidomics may contribute to this (Kim et al. 2021).

6.6 Mitigating dual-use risks of general-purpose viral engineering approaches

The scientific community needs to safeguard general-purpose viral engineering approaches that may make the misuse of pathogens easier, including certain computational tools. The latter may enable successful gene therapy and thus need to be guarded from misuse (Wec et al. 2021). In the near term, software developers should keep their tools specific to low-risk AAV until best practice norms for ethical and security-considerate use of computational tools have been developed. This is because sharing code and tools applicable to viruses capable of unassisted human-to-human transmission may inadvertently and irreversibly make their enhancement easier. One important strategy to mitigate misuse of general-purpose computational tools may be structured access solutions, as discussed in 4.4. Stakeholder discussions, including data and code sharing platforms like GitHub, are needed to identify balanced measures to reduce risks.

6.7 Lessons from assessing dual-use risks of viral vectors for gene therapy

There are multiple takeaway messages from analysing the dual-use potential of viral vectors for gene therapy. First, synthetic virology research for gene therapy already features dual-use risks and such risks will only increase. I identify a drive towards universally applicable approaches and general-purpose methods that, if advanced in a more sophisticated form, will feature increasing dual-use risks. For instance, viral vector engineering drives the development of universal genetic elements for enhancing viral properties applicable to a broad range of viruses. Such universal genetic enhancement approaches deserve special dual-use oversight. Second, virus-specific work on AAV and other non-autonomously transmissible

viruses may be a low-risk alternative to more concerning viral vector work. Prioritising promising AAV work for gene therapy could be a way to maximise benefits while minimising dual-use risks. Lastly, viral vector engineering for gene therapy and cancer immunotherapy leads to advances in cutting-edge general-purpose viral engineering tools. In particular, computational tools may substantially increase the power and accessibility of viral engineering and thus deserve special dual-use consideration, which is the subject of the next chapter. .

Chapter 7: Artificial intelligence, large language models, and biodesign tools[17]

7.1 Introduction

Artificial intelligence (AI) has the potential to catalyse advances across many different areas of the life sciences. For example, AI tools are already driving protein design capabilities (Eisenstein 2023), antibiotic discovery (Liu et al. 2023), and the ability to understand the human genome (Dalla-Torre et al. 2023). In the longer term, AI may transform the nature of life sciences research through automated research capabilities (Boiko, MacKnight, and Gomes 2023). These developments will also strengthen health security, for instance, by empowering the ability to detect and respond to infectious disease outbreaks (Wang et al. 2021). However, as AI transforms the life sciences, this may also increase biosecurity risks through empowering the weaponisation and misuse of biological agents, including pandemic pathogens.

I differentiate between two forms of AI which, in different ways, exacerbate the risk of biological misuse: large language models (LLMs) and biodesign tools. Drawing on evidence from historical biological weapons programs, I analyse how these two different classes of general-purpose AI tools impact barriers to weaponising biological agents. Given advances in general-purpose AI systems have broad and fundamental impacts on the accessibility and modification of biological agents, this chapter focuses less explicitly on pandemic pathogens. I argue that LLMs and biodesign tools feature significantly different properties and risk profiles (see Table 7.1), which has important implications for appropriate risk mitigation strategies.

[17] This section has with minor modifications been turned into the following preprint: Sandbrink, Jonas B. 'Artificial intelligence and biological misuse: Differentiating risks of language models and biodesign tools.' *arXiv,* 2023. doi:10.48550/arXiv.2306.13952

AI applications may also increase biosecurity risks through indirect avenues. For instance, LLMs could also exacerbate misinformation and disinformation challenges (Goldstein et al. 2023), which could negatively impact the response and attribution of a biological event. Furthermore, LLMs might be misused as tools to radicalise and recruit or to coerce and manipulate scientists to acquire technical expertise for biological weapons development. While disinformation and manipulation risks need to be examined and addressed, these risks are less unique to biosecurity and are not the focus of this piece.

AI tools may exacerbate biological risks more drastically than risks of chemical misuse. Undoubtedly, AI has the potential to lower barriers to chemical weapons. Indeed, Urbina *et al.* have illustrated how an AI-powered drug discovery tool could be used to generate blueprints for plausible novel toxic chemicals (Urbina et al. 2022). However, increases in the potential harm of chemical weapons will be limited because of their non-transmissible nature. In contrast, as the following sections argue, AI may not only increase the accessibility of biological weapons but could also increase their ceiling of harm because it could enable the design of biological agents with unprecedented properties. Thus, the intersection between AI and the biological sciences has particularly pronounced security implications.

Table 7.1: Summary of characteristics, risks, and risk mitigation options for LLMs and biodesign tools

	Large language models (LLMs)	**Biodesign tools**
Definition	Tools trained primarily on natural language which can provide scientific information, access relevant online resources and tools, or instruct research.	Tools trained on biological data that are used for designing new proteins or other biological agents.
Examples	● Foundation models (e.g. GPT-4/ChatGPT) ● Language models optimised for assisting scientific work (e.g. BioGPT) ● Language model-based tools for autonomous scientific research	● ProteinMPNN, RFdiffusion ● Protein language models trained on genetic sequences (e.g. ProGen2) ● Smaller and more specialised tools (e.g. Ogden et al 2019)

	(e.g. Boiko et al. 2023)	
Developers	• Foundation models are developed by a limited number of well-resourced companies. • Science-specific models or applications of foundation models are developed by more distributed creators.	• The majority of biodesign tools are developed in a very distributed and open-source manner. • A small number of large-scale models have been developed by well-resourced companies.
Major risks	Lower barriers to accessing and misusing biological agents across the whole spectrum of actors (i.e. individuals, groups, and state programs): • Providing information on dual-use topics • Providing lab assistance and, eventually, autonomous research • Identifying avenues for misuse • Creating a perception of increased accessibility In the future, autonomous science tools may also increase the ceiling of capabilities.	Increased ceiling of capabilities, in particular for most sophisticated actors (i.e. state programs and well-resourced, sophisticated groups): • Enabling creation and misuse of pathogens much worse than anything known today • Enabling biological weapons attractive to state actors, e.g. targeted to populations or geographies In the short term, enabling the creation of hazardous proteins that are not picked up by existing gene synthesis screening.

Risk mitigation	• Only release powerful LLMs with structured access; consider limits on open sourcing and strengthen information security to prevent leaks
	• Pre-release evaluations by third parties and post-release reporting of hazards for foundation models to ensure developers have eliminated dangerous capabilities
	• Provide differentiated access to dual-use AI tools for science based on authentication of users
		• Review of dual-use risks before, during, and after biodesign tools development
		• For general-purpose biodesign tools, consider moving away from open access to structured access; this would enable future governance measures like differentiated access to certain capabilities based on authentication of users
		• Universal gene synthesis screening and development of screening based on functional prediction
		• Make biological weapons less attractive to well-resourced actors
		  ○ Strengthen intelligence and law enforcement
		  ○ Strengthen norms against biological weapons
		• Consider red lines for development of certain dangerous capabilities

7.2 Risks from large language models (LLMs)

7.2.1 Overview of large language models and related advances

The first class of AI tools that might enable misuse of biology are large language models (LLMs) that have been trained on scientific documents and discussion forums.[18] These are representative of a larger set of "AI assistants", which feature a broad spectrum of general-purpose capabilities and natural language outputs and inputs. LLMs can help with providing information, accessing relevant online resources and tools, and instructing research. Relevant tools include foundation models (e.g. GPT-4/ChatGPT) or language models optimised for assisting scientific work (e.g. BioGPT) (OpenAI 2023; R. Luo et al. 2022). One subcategory of the latter may be models with autonomous research capability, which includes the use of laboratory robots (Dama et al. 2023; Boiko, MacKnight, and Gomes 2023).

Different categories of models are developed by different groups of creators. Foundation models are products of large and expensive training runs and thus are currently developed by a small number of companies (Yang et al. 2023). The majority of cutting-edge foundation models have not been open-sourced and can be accessed through web interfaces and application programming interfaces (APIs). Scientifically-focused models or foundation model-based applications for autonomous research are developed in a somewhat more distributed manner. While BioGPT was developed by Microsoft and also required costly computational training, some autonomous science applications have been developed by more resource-constrained academic researchers (R. Luo et al. 2022; Boiko, MacKnight, and Gomes 2023; Dama et al. 2023). All of these science-focused tools are open source.

[18] Language models trained on genetic sequences, which predominantly output genetic sequences, are considered biodesign tools and not LLMs.

In the following, I discuss key ways in which LLMs might impact the risks of biological misuse. A key theme is that LLMs increase the accessibility to existing knowledge and capabilities, and thus may lower the barriers to biological misuse (see Figure 7.1b).

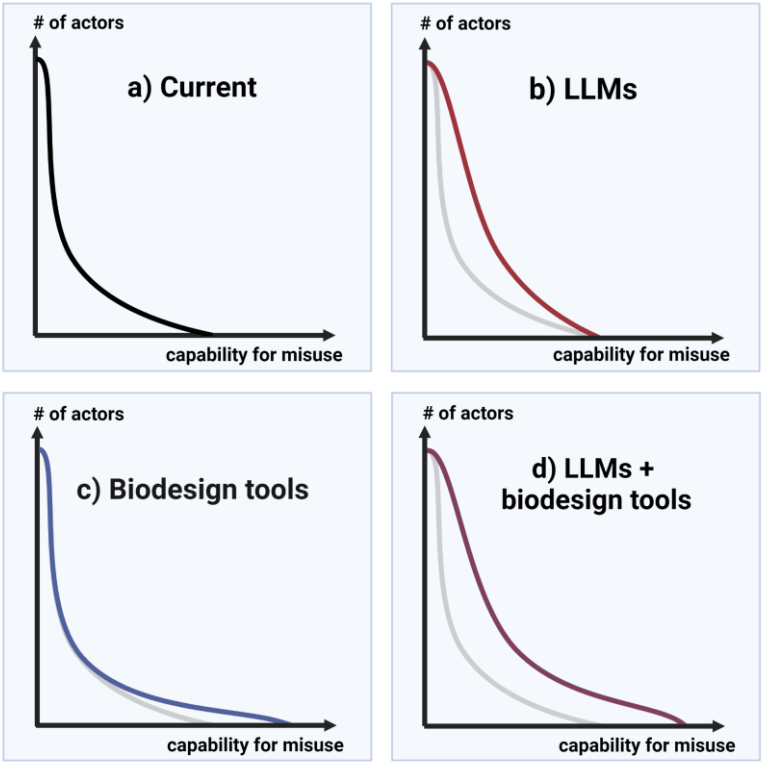


Figure 7.1: Schematic of effects on LLMs and biodesign tools on capabilities for biological misuse

Illustrative schematic of how artificial intelligence tools impact capabilities across the spectrum of actors with the potential to misuse biology. a) Currently most individuals are not able to access biological agents, and only a small number of actors are capable of causing large-scale harm. b) Large language models (LLMs) will increase capabilities across the spectrum of actors but are less likely to substantially raise the ceiling of capabilities. c) Biodesign tools will increase the ceiling of capabilities. d) The combination of LLMs and biodesign tools will increase the ceiling of capabilities and make such capabilities accessible to a significant number of individuals.

7.2.2 Teaching about dual-use topics

First, LLMs will enable efficient learning about dual-use knowledge. In contrast to internet search engines, LLMs can answer high-level and specific questions relevant to biological weapons development, can draw across and combine sources, and can relay the information in a way that builds on the existing knowledge

of the user. This could enable smaller biological weapons efforts to overcome key bottlenecks. For instance, one hypothesised factor for the failed bioweapons efforts of the Japanese doomsday cult Aum Shinrikyo is that it's lead scientist Seichii Endo, a PhD virologist, failed to appreciate the difference between the bacterium Clostridium botulinum and the deadly botulinum toxin it produces (Ouagrham-Gormley 2014, 142). ChatGPT readily outlines the importance of "harvesting and separation" of toxin-containing supernatant from cells and further steps for concentration, purification, and formulation. Similarly, LLMs might have helped Al-Qaeda's lead scientist Rauf Ahmed, a microbiologist specialising in food production, to learn about anthrax and other promising bioweapons agents, or they could have instructed Iraq's bioweapons researchers on how to successfully turn its liquid anthrax into a more dangerous powdered form (Ouagrham-Gormley 2014; Warrick 2006). It remains an open question how much LLMs are actually better than internet search engines at teaching about dual-use topics.

7.2.3 Identifying specific avenues to biological misuse

Second, LLMs can help with the ideation and planning of how to attain, modify, and disseminate biological agents. Already now, LLMs are able to identify how existing supply chains can be exploited to illicitly acquire biological agents. In a recent one-hour exercise, LLMs enabled non-scientist students to identify four potential pandemic pathogens and how they can be synthesised, which companies supply synthetic DNA without screening customers and orders, and the potential to engage contract service providers for relevant laboratory work (Soice et al. 2023). In the longer term, LLMs could also generate ideas for how to design biological agents tailored for a specific goal, such as what molecular targets would be best suited to produce a particular pathology.

7.3.3 Step-by-step instructions and trouble-shooting experiments

Additionally, LLMs could become very effective laboratory assistants which can provide step-by-step instructions for experiments and guidance for troubleshooting experiments. Such AI lab assistants will have many beneficial applications for helping less experienced researchers and replicating experimental methods from publications. However, these AI lab assistants might also support laboratory work for malicious purposes. For instance, a key reason for Aum Shinrikyo's failure to weaponise anthrax was that Seiichi Endo did not succeed at turning a benign vaccine strain of the bacterium into its pathogenic form, despite access to relevant protocols for plasmid insertion. Endo might have succeeded with an AI lab assistant to provide tailored instructions and help with troubleshooting. One crucial open question is how much of an additional barrier "tacit knowledge" plays, which, as discussed in Chapter 2, describes knowledge that cannot easily be put into words, such as how to hold a pipette or recognise when cells look ready for the next step of laboratory work. However, what is clear is that if AI lab assistants create the perception that performing a laboratory feat is more achievable, more groups and individuals might try their hand - which increases the risk that one of them actually succeeds.

7.2.4. Autonomous science capability

In the longer term, as LLMs and related AI tools improve their ability to do scientific work with minimal human input, this could potentially transform barriers to biological weapons. Firstly, LLMs can instruct laboratory robots based on natural language commands, which will make them easier to use (Inagaki et al. 2023). Secondly, LLMs can serve as the basis for autonomous science agents, which break tasks into manageable pieces, interface with relevant specialised computational tools, and instruct laboratory robots (Boiko, MacKnight, and Gomes 2023). Challenges relating to coordinating large teams under secrecy limited the Soviet and Iraq bioweapons programs and likely has also served as a barrier for terrorist groups

(Ouagrham-Gormley 2014). If autonomous science capabilities enable individuals and small groups to achieve large-scale scientific work, this will likely empower covert bioweapons programs.

7.3 Risks from biodesign tools

7.3.1 Overview of biological models with a focus on protein design tools

The second class of AI tools that might pose a risk of misuse are biodesign tools. I define biodesign tools as follows:

> ***Biodesign tools:*** *Artificial intelligence models that are trained on biological data that can help design new proteins or other biological agents.*

In the following risk assessment, I focus in particular on tools for protein design, which can predict viable protein sequences that fulfil specific functional characteristics.[19]. Examples of current tools meeting this definition include ProteinMPNN, RFDiffusion, ProGen2, and Ankh (Dauparas et al. 2022; Watson et al. 2022; Madani et al. 2023; Elnaggar et al. 2023). ProGen and Ankh are language models trained on genetic sequences - as for language models of this kind the output is a designed genetic sequence, these "genetic text"-based protein language models (PLMs) feature the risk profile of biodesign tools. Currently, biological design capabilities are still limited to creating proteins with relatively simple, single functions. However, eventually, relevant tools likely will be able to create proteins, enzymes, and potentially eventually whole organisms optimised across different functions.

[19] An example might be the generation of a genetic sequence encoding a soluble, thermostable, and readily produced protein that binds to x receptor. The protein design challenge may also be referred to as "inverse protein folding".

Biodesign tools are advanced both by companies, academics, and industry-academic partnerships. Companies have developed a number of influential larger-scale models, such as ProGen2 (Salesforce Research), Ankh (Proteinea), and the protein folding tools ESM-2 (Meta AI) and AlphaFold2 (Google DeepMind). However, academic groups, most notably the Baker Lab at the University of Washington, are also creating cutting-edge protein design tools (Dauparas et al. 2022; Watson et al. 2022). All of these tools have been published open source.

While this assessment focuses on larger, more general-purpose tools for protein design, the identified lessons may also translate to potentially smaller, more specialised tools like computational tools for predicting zoonotic potential (discussed in 4.4) and for optimising viral vectors (6.4). Such tools are developed by many different academic groups and companies based on more specialised training data for solving specific challenges. Next to tools for protein or organism design, there are also other machine learning tools with related dual-use implications, such as tools that shed light on host-pathogen interactions through predicting properties like immune evasion (Thadani et al. 2023) or through advancing functional understanding of the human genome (Dalla-Torre et al. 2023).

In the following, I examine key ways in which biodesign tools might impact on risks of biological misuse. In contrast to LLMs which mainly increase the accessibility of biological weapons, biodesign tools may increase the ceiling of capabilities and thus the ceiling of harm posed by biological weapons (see Figure 7.1c).

7.3.2 Sophisticated groups and increased worst-case scenario risks

First, as biodesign tools advance the ceiling of biological design, this will likely increase the ceiling of harm that biological misuse could cause. AI tools could enable sophisticated ways to optimise pathogens across

multiple properties, including transmissibility, virulence, and immune evasion. If released accidentally or deliberately, such pathogens might pose great catastrophic potential. It has been hypothesised that for evolutionary reasons naturally emerging pathogens feature a trade-off between transmissibility and virulence (Alizon et al. 2009). Biodesign tools might generate design capabilities that are able to overcome this trade-off. Thus, for the first time, humanity might face a security threat from pathogens substantially worse than anything nature might create, including pathogens capable of posing an existential threat.

There are a number of historical examples of ideologically extremist groups that might attempt to create such pathogens because they are motivated to cause maximal and indiscriminate harm (Torres 2018). One example is Aum Shinrikyo, the Japanese doomsday cult described in Chapter 1 which attempted to weaponise biology and successfully performed deadly sarin attacks in the Tokyo subway in the 1990s (Danzig et al. 2012). Through biodesign tools, such groups might be able to start a pandemic optimised for its catastrophic impact, which could even pose an existential threat to human society.

Bioterrorism with pathogens designed to cause maximal harm is a low-probability scenario; very few people have relevant motivations, and - even with AI tools - designing an optimised pathogen will constitute novel research involving multiple design-build-test iterations, which will require significant skills, time, and resources. Effective use of biodesign tools currently still requires both molecular biology and computational skills and well-informed target selection. However, these barriers to using biodesign tools may decrease with advances in large language models or other AI lab assistants.

7.3.3 State actors and new capabilities

Second, biodesign tools may be a key contributor to raising biological engineering capabilities in a way that makes biological weapons more attractive for state actors. Generally, state actors are only interested in

weapons systems that are predictable and do not hurt their own forces. Biological weapons have not met these criteria very well in the past. The United States never included bioweapons developed during the 1960s in its war plans due to their short shelf life and limited frontline usability due to the risk of harming friendly troops (Ouagrham-Gormley 2014). Iraq never deployed its bioweapons, likely because of a lack of certainty around its effectiveness and fear of retaliatory measures (Ouagrham-Gormley 2014; J. Miller, Engelberg, and Broad 2002).

If AI tools push the ceiling of biological design to make biological agents more predictable and targetable, this could increase the attractiveness of biological weapons. An example would be the engineering of pathogens to only spread in certain geographic areas or populations, which would reduce the risk of blowback on the own population. New biological capabilities may emerge relatively surprisingly; an example is that no one predicted the emergence of gene drive technology when CRISPR/Cas9 was first explored as a gene editing system. Well-intentioned researchers will likely be involved in demonstrating proof-of-concept for relevant capabilities as they emerge, which might create additional attention and experimental evidence.

7.3.4 Circumventing sequence-based biosecurity measures

Lastly, in the near term, biodesign tools may challenge existing measures to control access to dangerous agents based on their taxonomy and genetic sequence. Important mechanisms to prevent illicit access to toxins and pathogens are export controls for lists of agents, such as the Australia Group List, and the screening of organism sequences by gene synthesis providers (The Australia Group 2020; Diggans and Leproust 2019). Biodesign tools will make it easier to design new agents with a specific function that do not map onto existing taxonomic or functional categories and do not contain known sequences of concerning agents. Indeed, biodesign tools may enable the "recoding" of existing proteins or organisms by

finding substantially different sequences encoding for similar structures and functions - for instance for encoding the function of a known toxin. Already, researchers are recoding entire bacterial genomes (Fredens et al. 2019) and protein design tools such as RFdiffusion have started to disseminate such capabilities for proteins with relatively simple functions (Watson et al. 2022). Thus, taxonomy or sequence similarity-based controls will not be sufficient to prevent access to harmful biological agents in an age of AI-powered biological design.

7.4 Mitigating biosecurity risks from large language models

The properties of LLMs and biodesign tools and their divergent risk profiles have important implications for risk mitigation. For LLMs, risks require urgent and substantial action. Individuals working at cutting-edge AI companies have highlighted the potential for LLMs to exhibit "dangerous capabilities", which includes the proliferation of weapons, and have advocated for appropriate risk mitigation (Shevlane et al. 2023). Existing LLMs may already make it easier for a technical expert in one area to skill up in complementary areas required for biological weapons development. While questions remain around how much LLMs will make dual-use information more accessible and will lower tacit knowledge barriers, waiting with risk-mitigating interventions until more evidence has accumulated is not a viable strategy. Fast and unpredictable advancements in LLM technology necessitate urgent governance measures. Importantly, it is very difficult to predict the capabilities of an LLM before its training and fine-tuning (Wei et al. 2022). If a powerful foundation model does not seem to feature dangerous capabilities and is open-sourced, someone may later finetune the foundation model so that it develops the capability to instruct biological misuse. In this case, it would be too late to retract access to the model. Additionally, as training runs take a long time and require significant resources, the tracks for the next generation of LLMs are laid many months before their release.

7.4.1 Controlling access to dual-use capabilities

One key way in which risks from LLMs may be mitigated is by controlling who and for what purposes accesses certain dual-use capabilities. Access controls might be effective and feasible because of how LLMs cause biosecurity risks and who is developing them: First, LLMs lower barriers to biological misuse for moderately resourced and skilled groups and individuals. Access control may be well-suited to reducing risks from such less sophisticated actors, who may be more opportunistic and are unlikely to be able to circumvent access controls. Second, because a smaller number of developers are able to train expensive cutting-edge foundation models, it may be more feasible to implement access controls. However, widely disseminated open-source alternatives are currently often only a few months behind cutting-edge models, especially as new low-cost methods to train LLMs are identified (Patel and Ahmad 2023; Hu et al. 2021).

Access controls rely on the fact that models are not openly disseminated. Open dissemination prevents the mitigation of any kind of misapplication of LLMs, including the generation of harmful content like hate speech. Therefore, 'structured access' has emerged as a new paradigm for the safe deployment of foundation models (Shevlane 2022). Different from open access release which enables users to download the model and change its code, structured access is mediated through a web interface and an application programming interface (API)[20]. These interfaces can limit applications of models for unintended purposes and ensure safety standards of applications built on the model (Shevlane 2022). Thus, the key to effective mitigation of biosecurity risks from LLMs is that such models are not open-sourced and that developers adopt good information security procedures to prevent the leak of model weights.

[20] As discussed in 4.4, APIs are the interface through which different pieces of software interact with each other.

7.4.2 Pre-release model evaluations

Biosecurity risks from foundation models can be mitigated by ensuring that they do not feature dangerous capabilities at release. Restricting dangerous capabilities for foundation models may be warranted if the benefits for risk reduction are significant and the cost for beneficial applications is small (Anderljung and Hazell 2023). This may be the case for dangerous information relevant to biosecurity. Foundation models accessed by the general public do not need to be able to brainstorm ideas for misuse or to instruct dual-use scientific experiments. A crucial question will be to define where to draw the line for what biosecurity-relevant information foundation models should not disclose.

Leading companies and AI governance scholars are coalescing around pre-release evaluations as a key tool for preventing the release of models with dangerous capabilities (Shevlane et al. 2023). Such pre-release evaluations were for instance used for GPT-4 (OpenAI 2023). Pre-release evaluations involve an external audit of foundation models with a set of tests and questions, which should also include explicit evaluation of dangerous biological capabilities. Such biosecurity evaluations could include queries related to different steps of biological weapons development and viral synthesis, however, details should likely not be public to prevent targeted circumvention of queries. Mandating pre-release evaluations would incentivise developers to reduce dangerous capabilities throughout the development and release of their model. This could involve removing certain dual-use scientific documents from training data and using reinforcement learning from human feedback[21] (RLHF) to finetune models not to disclose harmful information (Bai et al. 2022). Additionally, developers could embed classifiers to screen queries and outputs of LLMs for harmful content, as well as mechanisms to track suspicious activity across a series of queries.

[21] Reinforcement learning from human feedback describes the practice of training a model to perform more in a way that is desirable to human users. For instance, after initial training, large language models are being exposed to human users who rate different model outputs; these ratings are then used to train the model to give answers that are more desirable.

7.4.3 Differentiated access to science tools

For LLM-based tools for scientific research, access controls would need to be much more differentiated. Many users of LLM-based AI lab assistants or autonomous science tools have legitimate reasons to access dual-use capabilities. For example, the synthesis of an influenza virus from its genome is crucial for influenza research to improve vaccines, therapeutics, and diagnostics. Thus, the dual-use capabilities of AI lab assistants need to be governed using differentiated access controls. This resembles the challenge faced by gene synthesis providers to prevent illicit access to synthetic DNA while avoiding friction for legitimate users. Gene synthesis screening generally features a combination of customer and sequence screening (U.S. Department of Health and Human Services 2010). Similarly, AI lab assistants could require users to authenticate their identity and to allow concrete queries relating to controlled agents and experiments based on documentation of biosafety and biosecurity review. An important challenge will be to ensure equitable access across the globe, including to researchers working outside of well-recognised universities.

7.5 Mitigating biosecurity risks of biodesign tools

Risks from biodesign tools require monitoring as they may be significant but are still mostly on the horizon. Biodesign tools might in particular cause biosecurity risks through pushing the ceiling of biological design. This is only just starting to take place, so risks are mostly still at the horizon and only ill-defined. There is a diverse set of AI-empowered tools that advances biological design capabilities, and it is difficult to predict how advances will take place. For instance, it is unclear whether large models trained only on biological sequence data will lead to significant advances in design capabilities, or whether more functional data and more tailored tools will be required to lead to significant advances. Additionally, development of biodesign tools may take place in a way that is more dispersed and more gradual than that

of LLMs, and academia may play a significant role in developing cutting-edge tools. Governments need to establish forums in which they can monitor risks and can create nimble governance strategies. These measures should be informed by biosecurity experts and tool developers, who should be required to practise dual-use review before, during, and after biodesign tools development.

7.5.1 Deterring sophisticated actors

Biodesign tools may in particular increase offensive capabilities for the most technically advanced actors, which has important implications for risk mitigation. By definition, biodesign tools will in particular advance capabilities to engineer and enhance biological agents - which by definition would constitute novel research and thus require significant skill and resources. Misuse by relevantly skilled and well-resourced actors will be difficult to prevent using access controls, because they either have access to relevant tools because of relevant legitimate work or they have the technical capabilities or network to circumvent access restrictions. Generally, other interventions may be more promising to stop technically advanced actors. For sophisticated non-state actors, it may be most promising to strengthen intelligence and law enforcement to detect and stop instances of misuse. Sophisticated attempts will involve design-build-test iterations which will leave intelligence signatures. For state actors, it may be most effective to prevent them from developing biological weapons by ensuring biological weapons stay unattractive. This might be achieved by strengthening norms against any form of biological weapons (including ones that do not harm humans), advancing a verification regime for the Biological Weapons Convention, and developing robust methods to attribute and detect biological attacks. More generally, technology developers and the international community could consider the formulation of red lines to avoid the development of capabilities that might be considered too dangerous to global security.

7.5.2 Structured access for biodesign tools

Open-source publishing of computational tools for biology is valuable, however, it may nevertheless be important to explore the use of structured access for a subset of biodesign tools with biosecurity risks (Smith and Sandbrink 2022). One reason is that LLMs will likely make biodesign tools more accessible, which might make the ceiling of capabilities much more accessible (see Figure 7.1d). For instance, LLMs could provide natural language interfaces to using biodesign tools and AI lab assistants might help with turning biological designs into physical agents. Additionally, structured access methods for biodesign tools might provide a baseline through which risks can be monitored and eventually governed. If uninhibited open-source development of general-purpose biodesign tools becomes the norm, it may be much more difficult to govern risks in the future. Lastly, structured access methods may also be important to tackle the shorter-term risk that biodesign tools are used to evade biosecurity screening. For instance, structured access could be required for protein design tools that are able to create functional equivalents of controlled toxins and pathogens, such as agents on the US Federal Select Agent Program or the Australia Group export control lists (Centers for Disease Control and Prevention and U. S. Department of Agriculture 2022; The Australia Group 2020). Pre-release capabilities evaluations could also help identify relevant tools that should only be released under structured access. Providers of relevant tools might for a subset of dangerous applications require authentication of users and documentation of biosafety and dual-use review. Many biodesign tools developers might not have the resources to maintain access to their tool through a web interface or API, so a publicly maintained credentialled access platform might be needed.

7.5.3 Gene synthesis screening

Lastly, advances in LLMs and biodesign tools increase the importance of biosecurity measures at the transition from the digital to the physical. This includes measures to stop illicit access to synthetic DNA and other relevant synthetic biology services (see Chapter 3). Thus, advances in AI renew the urgency for

governments to create a mandatory baseline of gene synthesis screening. Additionally, advances in biological design also necessitate in-step advances of screening methods. For example, it may be possible for future synthesis screening tools to predict the function of novel sequences. To this end, AI developers, biosecurity experts, and companies providing synthesis products could collaborate to develop appropriate screening tools. As countries introduce gene synthesis screening regulations, these policies need to consider the need to expand screening to functional equivalents of controlled agents and eventually to completely novel hazardous agents.

7.6 Conclusions on artificial intelligence and biosecurity risks

It is yet uncertain how and to what extent advances in artificial intelligence will exacerbate biosecurity risks. Because of the rapid advances in artificial intelligence and its transformative potential, it is important to create an understanding of likely future developments and appropriate governance options to enable a timely policy response. To this end, it can prove useful to differentiate between LLMs and biodesign tools and possible differences in impacts and governance options. This distinction might blur as different tools are combined and tools are developed that leverage both natural language and biological data.

Risks at the intersection of AI and biosecurity may have policy implications that go beyond their immediate mitigation. If AI makes the misuse of biology more accessible, this strengthens the need for mitigating dual-use risks in the life sciences more generally. At the same time, biosecurity risks are a concrete instantiation of a broader set of artificial intelligence risks that could catalyse general AI governance measures.

To create the best evidence for appropriate risk mitigation, work to answer crucial open questions is needed. This includes monitoring how new AI tools are integrated into life sciences research processes,

assessing how AI is regulated more generally and ensuring representation of biosecurity-specific considerations, and exploring how key governance measures such as dangerous capability evaluations could be realised. If risks from AI can be effectively mitigated, this sets the groundwork for enabling AI to realise its very positive implications for the life sciences and human health.

Conclusions from the case studies

Across the above case studies, I identify a set of dual-use risks that may be significant but have not attracted close attention previously. First, wildlife virus discovery may lead to the identification of new potential pandemic viruses - and thus features misuse risks that are similar to ePPP research. Second, viral vector research for vaccines and gene therapy may create insights into viral enhancement which may be transferable to related pathogens. One particularly concerning subset may be universal genetic elements for immune modulation, which are applicable to many pathogens. Third, viral vector work drives general-purpose methods for viral engineering, including computational tools. Fourth, artificial intelligence advances may lower barriers to the misuse of pandemic pathogens and could increase the ceiling of pathogen enhancement and design. These dual-use risks may be on a scale similar to risks previously discussed, for instance as part of debates on the enhancement of potential pandemic pathogens (Duprex et al. 2015). However, these risks are associated with a broader set of research and technology development, much of which is currently not subject to dual-use assessment or oversight. Thus, the consideration of dual-use potential needs to be expanded beyond research traditionally associated with dual-use risks, such as microbiology, to also capture new relevant research, such as viral vector engineering for gene therapies and vaccines. The ambition to develop next generation medicines like gene therapies increasingly drives viral engineering. Accordingly, risk mitigation strategies need to adapt.

Across the four case studies, I have considered how risks may be addressed and propose concrete strategies. For wildlife virus discovery and characterisation work, many risk mitigation measures are possible. However, in some cases, risks may exceed benefits. Certain research may create knowledge that may do more harm than good. Thus, projects need to be assessed for whether risks outweigh benefits. In contrast, for viral vector engineering for vaccines and gene therapy, risk mitigation frequently comes down to prioritising low-risk avenues to solving a given challenge. Many low-risk approaches are very promising, for instance, RNA vaccines and AAV-based gene therapies, and thus prioritising these low-risk alternatives may be associated with no loss of benefits. Lastly, the governance of general-purpose tools, in particular relating to artificial intelligence, is of increasing importance for mitigating biotechnology misuse. To prevent the indiscriminate proliferation of the ability to create pandemic-capable agents, the scientific community should consider practices for eliminating dangerous capabilities before the release of LLMs and by managing access to these tools.

The new areas of dual-use risks highlighted by these case studies showcase that risks are evolving alongside advances in science. Many of the debates and solutions of the past, discussed in Part I, need to be expanded to address dual-use risks arising from a much broader set of research not traditionally associated with dual-use risks. The first step in this process is identifying why existing dual-use oversight falls short in addressing these emerging risks, which is the subject of the next part of this thesis.

PART III: ANALYSIS

Chapter 8: Limitations of dual-use research of concern oversight

In Part 1 of this thesis, I described the historical developments that have led to existing dual-use oversight measures. In Part 2, I considered specific case studies of research areas relating to virology that have dual-use risks but are not considered by existing oversight. Now, in Part 3, I turn to analysing why existing dual-use management fails to capture these dual-use risks. In the first half of this chapter, I analyse US DURC oversight, which I argue exemplifies the prominent approach to dual-use risk mitigation. In the second half of this chapter, I draw on historical debates and my case studies to identify limitations of existing dual-use management. In Chapter 9, I then focus on analysing one particular limitation highlighted by my case studies, the failure to capture what I call "transfer risks". In Chapter 10, I analyse a crucial barrier to the mitigation of transfer risks: the focus of existing dual-use management on individual projects. My analysis of existing dual-use management approaches and their shortcomings forms the foundation for proposals to improve dual-use management in Part IV.

8.1 United States DURC oversight defines dual-use management

When evaluating the existing approach to dual-use management, I focus on the United States. US funding bodies and research institutions drive many advances in the life sciences and biotechnology, including research on the areas discussed in Chapters 4-6. Thus, US policies on dual-use research are particularly important for managing emerging dual-use risks. Furthermore, the framing and scope of the US policies are influential internationally and have shaped the global discourse on dual-use management. The US definition of dual-use research of concern (DURC) has been picked up in other national and international discussions (Salloch 2018; World Health Organisation 2021c). For instance, in 2013, the WHO hosted a consultation on "DURC", which not only used the US DURC language but also drew on the US list of

experiments of concern (World Health Organisation 2013). Thus, similar to the Brussels effect for other regulations (Bradford 2020), where strict EU rules shape global standards, one might describe a Washington effect for dual-use management.

Dual-use management in the United States is defined by DURC oversight. The 2012 and 2014 DURC policies established this oversight and have since defined how US research institutions assess and mitigate dual-use risks. The DURC policies are complemented by the Federal Select Agent Program (FSAP) and the Potential Pandemic Pathogen Care and Oversight policy (P3CO). As discussed in Chapter 2, FSAP provides regulations for the physical possession of particularly dangerous pathogens and toxins and P3CO oversees Department of Human Health and Services (HHS) research grants for the enhancement of potential pandemic pathogens. Nevertheless, the DURC policies constitute the foundation of dual-use risks review at US federal funding bodies and US-funded institutions. In March 2023, the US National Science Advisory Board on Biosecurity (NSABB) published a report which evaluated the effectiveness of the existing US biosecurity policies and provided recommendations for how to strengthen them (National Science Advisory Board for Biosecurity 2023). The leading recommendation was to create a unified biosecurity oversight framework, which integrates elements of the existing DURC and P3CO policies. Where relevant, I will touch on the NSABB recommendations throughout this and the following chapters.

The DURC policies rely on the definition of DURC as the subset of dual-use research with significant potential for misuse. As discussed in Chapter 1, NSABB coined the concept of DURC in 2007 (National Science Advisory Board for Biosecurity 2007). DURC is defined as:

> ***"Dual-use research of concern (DURC)*** *is life sciences research that, based on current understanding, can be reasonably anticipated to provide knowledge, information, products, or technologies that could be directly misapplied to pose a significant threat with broad potential*

> *consequences to public health and safety, agricultural crops and other plants, animals, the environment, materiel, or national security."* - 2012 DURC policy (U.S. Department of Health and Human Services 2012)

The 2012 DURC policy established a review for DURC at federal institutions that conduct or fund life sciences research (U.S. Department of Health and Human Services 2012). Despite featuring the above broader definition of DURC, the scope of review was limited to the 7 classes of experiments identified initially in the Fink Report and 15 select agents and toxins from the Federal Select Agent Program (FSAP) list of access-controlled agents (see Tables 1.1 and 1.2 of Chapter 1). The policy states that if relevant research was identified to constitute DURC, risks and benefits should be assessed and risk mitigation approaches considered. Example risk mitigation approaches include the modification of experiments, special consideration around the venue and mode of communicating research results, and education on dual-use risks.

The 2014 DURC policy, sometimes referred to as the institutional DURC policy or iDURC, extended review of research for DURC to all US federally funded institutions and all US institutions conducting work on the agents and experiments captured by the policy (U.S. Department of Health and Human Services 2014). The policy establishes responsibilities for principal investigators and institutional review entities to review research for DURC, assess risks and benefits, and report resulting risk mitigation plans. This review process was limited to the same 15 select agents and toxins and 7 categories of experiments.

As discussed in Chapters 1 and 2, the DURC policies were also established to collect information for future updates of dual-use oversight, a process which has now yielded the March 2023 NSABB recommendations. Furthermore, despite the narrow approach to DURC oversight set out by these

policies, some scientists and institutions proactively review a greater range of research (S. W. Evans et al. 2021).

8.2 Characteristics and limitations of DURC oversight

8.2.1 List-based review

DURC policies require review of select experiments on high-consequence pathogens and toxins in order to mitigate dual-use risks where appropriate. These categories of experiments and select agents are defined by exhaustive lists (see Tables 1.1 and 1.2 in Chapter 1). Thus, under these policies only these experiments require oversight, and any additional dual-use oversight is voluntary. The new NSABB recommendations for updating the US biosecurity policies includes the suggestion to no longer limit review to a list of select agents but instead all "research that directly involves any human, animal, or plant pathogen, toxin, or agent that is reasonably anticipated to result in one or more of the seven experimental effects" (National Science Advisory Board for Biosecurity 2023). This is in line with the past debates and my own findings that I present in this section.

List-based approaches for governing life sciences research feature limitations. As life sciences advance, lists may become quickly outdated. The limits of list-based DURC oversight were recognised during the gain-of-function (GOF) debate discussed in Chapter 1, a debate which started one year before the first US DURC policy was even published. The GOF debate highlighted that DURC oversight limited to select pathogens would not capture all research on high-risk pathogens with potential for accidents or direct misuse. While work on avian influenza would be covered, gain-of-function experiments on high-risk coronaviruses or yet unknown pathogens would not be. Furthermore, experts from microbiology, public health, and epidemiology questioned whether, given the non-trivial risks of accidental releases, such experiments should be conducted at all and called for a more detailed review of risks and benefits at the

funding stage - a practice not adequately covered through the DURC policies, which mainly focus on risk mitigation and not whether to conduct an experiment at all (Duprex et al. 2015). In response to these gaps, the P3CO framework was created in 2017 to enforce funding stage review of all research involving the enhancement of potential pandemic pathogens.

My case studies also highlight that list-based DURC oversight may struggle to capture newly discovered or designed agents. For instance, DURC oversight specific to a list of select agents invariably fails to capture research to discover new high-consequence viruses. I discussed in Chapter 4 the dual-use potential of large-scale efforts to characterise viruses collected from nature. Results of such research could potentially credibly identify pandemic-capable viruses - agents that may be used to cause a global pandemic. Thus, viral characterisation experiments, such as testing viral binding affinity to human receptors, have dual-use potential if conducted on agents that may be transmissible in humans, as creating evidence for human transmissibility may be a crucial dual-use insight that could inspire misuse. Additionally, oversight focused on existing agents will be fundamentally challenged by the capability to design new biological agents, including with biodesign tools. For instance, soon it will likely be possible to create completely novel toxins with properties at least as concerning as select agents. This means that in the future dual-use policies will likely need to be based on agent properties rather than taxonomy, similar to P3CO oversight for potential pandemic pathogens.

Nevertheless, lists are a useful tool to enable easy and consistent implementation. Indeed, the creation of the DURC concept and the associated limitation of dual-use oversight to a manageable portion of research involving select agents has been hailed as a significant achievement for governance (Imperiale and Casadevall 2018). One reason for the list-based nature of the DURC policies might be the influence of the chemical, biological, radiological and nuclear (CBRN) defence community, which uses lists in export

controls, including for nuclear materials.[22] However, a list-based approach may feature intrinsic limitations when used for biology, where risks cannot be cleanly stratified along species boundaries.[23]

One way to retain the benefits of lists as a governance tool might be to use illustrative lists rather than exhaustive lists. Illustrative lists can make an underlying principle interpretable and enforceable. For instance, Israel's dual-use oversight law in principle is based on the US select agent list; however, this list is non-exhaustive as oversight also applies to work on other pathogens if risks are identified during an experiment (Lev 2019). Non-exhaustive or illustrative lists allow slightly more flexibility in interpretation, while the components of the list can still ensure a minimal baseline for what research is controlled. Because of the limitations of exhaustive lists seen in DURC oversight, P3CO was established by providing guiding principles only (see Box 1.3 in Chapter 1). However, P3CO could have potentially been made more effective by also featuring an illustrative list of example experiments.

8.2.2 Focus on microbiology

US DURC policies are not only limited by the exhaustive list of experiments and agents they apply to but also by the underlying principles that form the basis of the list. One such principle is a focus of oversight on microbiological experiments directly conducted on high-consequence pathogens and toxins. Review of microbiological experiments may be the easiest to implement as microbiologists are already used to biosafety and select agent regulations. However, this focus comes at the cost of limited flexibility regarding dual-use research that does not involve direct work on pathogens or toxins.

[22] Suggested by Claire Standley in private communication.

[23] Millett et al. (manuscript in peer review)

One example of dual-use work not captured by a focus on microbiology experiments is research on gene drives. As discussed in Chapter 1, gene drives are a genetic engineering approach which propagates a particular (modified) gene throughout a population. As gene drives consist of little more than genetic instructions encoded in a host genome to copy specific genetic code between genomes, gene drives do not resemble classical organisms. Nevertheless, they may spread exponentially and change whole animal populations, an ability with notable misuse potential.[24] Thus, advances in biotechnology mean that work with security implications might not resemble research on classical pathogens.

Furthermore, even dual-use insights on pathogen engineering increasingly come from areas beyond microbiology research. In Chapters 5 and 6, I analysed how research on viral vector vaccines, gene therapy, and oncolytic viral therapies could drive misuse potential. For instance, immune evasion methods or future attempts to enhance the transmissibility of transmissible vaccines may create unique insights into how to enhance viruses for causing global harm. If future revised DURC policies would extend review beyond select agents, this would mean that viral vector engineering for vaccines, gene therapy, or cancer research might also be captured - in this case, additional educational work would be needed to engage researchers in relevant areas that are not as familiar with dual-use risks.

Lastly, a focus on microbiology assumes that misuse will look like the enhancement of a pathogen for military purposes, such as the enhancement of lethality or usability. However, biotechnology features not just dual-use risks relevant to the deliberate release of pandemic viruses, but also to areas beyond the scope of this thesis like the manipulation or surveillance of populations. Neuroscience advances might create new dual-use insights on how to manipulate attitudes and beliefs (Mahfoud et al. 2018). At the same time,

[24] Misuse risks of gene drives are limited by their unique nature. Gene drives spread with each reproductive cycle and thus propagate relatively slowly. Importantly, gene drives may be reversed by a different gene drive.

increasingly abundant genomic sequencing may be used to identify and track individuals or ethnic groups. For example, alongside other surveillance methods, Chinese authorities have engaged in extensive DNA profiling of individuals of the Uyghur ethnic minority, which raised international human rights concerns in 2021 (Normile 2021). Thus, dual-use risks of advances in sequencing technologies and tools for analysing resulting data streams need to be taken seriously.

8.2.3 Focus on insights driving peak capabilities rather than accessibility

Current DURC oversight focuses on research that may raise "peak capabilities" rather than research that disseminates existing capabilities (see Box 8.1 for definitions).

New or raised peak capabilities often arise from cutting-edge research, and may initially be accessible to a small number of experts before slowly diffusing more broadly. For example, peak capabilities captured by existing DURC oversight include new pathogen enhancement methods or the ability to weaponise pathogens.

Box 8.1: Defining peak capabilities and accessibility

Research that raises peak capabilities: *Generating knowledge that enables people to do things (with dual-use potential) that no one could do before.*

Research that disseminates capabilities: *Research that generates (new) knowledge that enables more people to do things (with dual-use potential) that others (but usually many fewer individuals) could do before.*

However, misuse risk of biotechnology may not just depend on peak capabilities but also the accessibility or dissemination of a given capability. In particular, risk of misuse by less well-resourced groups or single individuals is proportional to the number of individuals with access to relevant dual-use capabilities: for

instance, the more people have access to the ability to start a pandemic, the greater is the chance of misuse. The ability to start a pandemic is becoming increasingly accessible given advances in molecular biology discussed in Chapter 3, including increasing accessibility of synthetic DNA and new molecular biology approaches. In the wake of COVID-19, many scientists shifted towards studying SARS-CoV-2 and acquired new synthetic virology skills (Pfeiffer and Dermody 2021; Musunuri et al. 2021). Given the trend of increasingly accessible dual-use biotechnology, dual-use management aimed at preventing misuse needs to capture accessibility-increasing research. Existing DURC oversight fails to do so, as demonstrated by the following examples.

First, DURC oversight does not capture work that reduces the tacit knowledge skill barrier to synthesise viruses. As introduced in Chapter 1, tacit knowledge describes certain uncodified knowledge which may be difficult to convey through language, such as the ability to ride a bike (MacKenzie and Spinardi 1995). A subset of tacit knowledge, so-called "weak tacit knowledge", may be rendered explicit through writing (H. Collins 2010); such knowledge may be critical for preventing the proliferation of biological weapons (Revill and Jefferson 2014). Examples of weak tacit knowledge also include specific technical protocols, including viral booting methods for the synthesis of infectious viruses from synthetic DNA.

A concrete example of work that reduced the tacit knowledge barrier is the protocol with step-by-step instructions on how to synthesise recombinant SARS-CoV-2 published by Xie *et al.* in early 2021 (Xie et al. 2021). The publication of such a detailed protocol for the booting of potential pandemic viruses may irreversibly reduce the need for specialised weak tacit knowledge and may enable the creation of pandemic-capable pathogens by a broader set of individuals with more basic laboratory skills (Pannu et al. 2021). Hence, protocols such as the one by Xie *et al.* should be subjected to dual-use review beyond the institutional review of the underlying experiments on potential pandemic pathogens. Without question,

it is critical to enable relevant researchers to create recombinant SARS-CoV-2 to promote the fast development of countermeasures. However, the upsides of sharing protocols need to be weighed against possible global security implications of publication. Alternative methods of sharing this protocol may have preserved the majority of its upside while limiting risks.

Large language models and AI lab assistants will likely be the development that will reduce tacit knowledge barriers the most over the coming years (see Chapter 7). These AI tools draw on protocols such as that by Xie et al. and make them even more accessible by advising on laboratory specific set-ups and providing help with troubleshooting. It is a crucial open question what kind of laboratory work public versions of these models should instruct. Arguably, the general public does not need to be able to receive detailed instructions and troubleshooting for synthesising influenza virus.

Second, DURC oversight fails to capture experiments and biotechnology that increase access to certain capabilities. One relevant past example is the synthesis of horsepox, which detailed a strategy to stitch together large poxviruses. As the authors of the horsepox synthesis paper note, the possibility to synthesise smallpox was already known before their work and also others would have been able to achieve such a feat (Noyce and Evans 2018). Despite this paper thus not generating a wholly new threat, it very likely increased the number of individuals capable of successfully creating smallpox from synthetic DNA.

My case studies highlight other advances that increase accessibility to certain capabilities. In Chapter 6, I discuss general-purpose viral engineering methods which not only increase peak capabilities for viral enhancement but also reduce the need for expert judgement in this process. This may include general-purpose massively parallelised directed evolution approaches or computational tool-guided viral engineering. Additionally, universal genetic elements for viral enhancement, such as the immune

modulating short-non-coding DNA fragments put forward by Chan *et al.* (Chan et al. 2021), may require less skills to be adapted to a new organism than virus-specific enhancement approaches. Thus, such plug-and-play methods for changing viral properties may enable a greater number of non-specialists to enhance viruses.

To comprehensively address biotechnology misuse, dual-use management needs to consider not only new peak capability-driving insights but also knowledge that increases the accessibility of certain dual-use capabilities. However, governing dual-use risks that increase accessibility may be more difficult than governing dual-use peak capabilities. Insights on peak capabilities can be more clearly identified; for instance, it is pretty clear that identifying genetic changes to turn avian influenza virus mammalian-transmissible is a new capability that features misuse potential. Thus, peak capabilities are relatively easy to circumscribe and regulate. In contrast, insights that increase accessibility arise more on a spectrum; some research contributes more to increasing accessibility, some less. Accessibility might thus advance more diffusely. Nevertheless, there may be specific dual-use insights that clearly significantly increase the accessibility of dangerous dual-use capabilities. For example, step-by-step protocols to create a potential pandemic pathogen are relatively straightforward to circumscribe and identify as featuring significant risks; and thus could be subject to oversight. Given dual-use capabilities are advanced not just by individual high-risk experiments but also many other projects diffusely creating relevant knowledge, new governance methods could be considered that go beyond a binary classification of DURC or not DURC. For instance, incentive systems could be used to make studies more unattractive based on where they are on the dual-use spectrum. I discuss such solutions in Part IV.

In the meantime, for large language models and related AI systems, it is likely that norms will emerge for what laboratory work public models should give detailed help with. When testing Claude2 shortly after

its launch in July, it refused many more queries relating to giving instructions for influenza virus synthesis than the existing chatGPT/GPT-4. Initially, such decisions will likely be based on the risk tolerance of model developers. Thus, government guidance could be helpful to create universal norms.

8.2.4 Focus on publicly funded research

Lastly, existing oversight mechanisms are limited in what research they capture. For instance, DURC oversight extends solely to US federally funded research. Privately funded research is not captured at all. However, dual-use research may increasingly come from biotechnology companies. For instance, my case studies showcase the potential for gene or cancer therapy companies to advance dual-use research. A related limitation is the lack of a global nature of dual-use oversight. Knowledge spreads globally and biological events have international reach. If strong oversight exists in one country, activities might simply move to a different location. Thus, exploring global approaches to dual-use oversight will be critical. The 2022 WHO guidance framework may help on this mission (World Health Organisation 2022b). The May 2023 NSABB recommendations also point at the need to expand the purview of existing policies. Not only do the recommendations suggest for dual-use oversight to apply to all relevant research on US soil and overseas research funded by US funding bodies, but also they also call for renewed efforts to strengthen and harmonise international norms on dual-use review (National Science Advisory Board for Biosecurity 2023).

8.2.5 Not capture any in silico work

The current DURC policies do not capture any *in silico* work, such as the creation of new designs for new biological agents. This fundamentally limits how these policies can mitigate dual-use potential of new artificial intelligence tools. At the moment, the line for what might qualify as DURC is drawn so that it only captures actual physical experiments. This makes sense as long as computational tools are not

powerful enough to eliminate the need for laboratory testing and iteration on created agents. However, once biodesign tools become powerful enough that they produce functional designs within just a handful of tries, this means that the importance of physical testing will fall away. The same may be true for computational abilities to predict pathogen properties, which could at some point with high confidence identify potential pandemic-capable pathogens from their genomes. Thus, over time, dual-use policies need to increasingly also consider the dual-use risks of outputs from *in silico* design and prediction processes.

8.2.6 Focus on research with direct misuse potential and individual projects

In the following two chapters, I evaluate two specific core characteristics of DURC oversight in greater detail. Chapter 9 discusses how the current DURC definition only captures research with "direct" misuse potential. Thus, research with dual-use insights transferable to pathogens is not subject to risk mitigation, research which includes many examples from my case studies. I identify characteristics of this neglected class of dual-use research which may enable risk mitigation.

In Chapter 10, I analyse the focus of DURC oversight and related risk mitigation strategies on individual projects. I define this as "individual project oversight" and propose how improvements to this approach could strengthen dual-use oversight.

Chapter 9: Transfer risks as a challenge for dual-use management

Currently, DURC oversight focuses on individual experiments that yield results with potentially substantial dual-use risks. However, actually individual high-risk experiments may only be a small part of what advances dual-use risks overall. Sociologists have argued that it is not individual publications of high-risk research that drives risk of misuse but broader scientific and technological developments. As discussed in Chapter 2, Kathleen Vogel's incremental change-based model of progress suggests that also misuse risks may be driven by broader scientific developments and not individual studies (Vogel 2012, 66). Furthermore, analysis of past barriers to bioweapons has highlighted that a crucial bottleneck is frequently local tacit knowledge related to the activity that is pursued; thus, considering the directionality of overall research trajectories and tacit knowledge development is more important than controlling individual high-risk studies (Ouagrham-Gormley 2014, 162). Additionally, advances in artificial intelligence will be able to synthesise and combine findings across a large number of studies to generate new related dual-use capabilities.

Thus, it is critical to consider the dual-use potential of incremental scientific progress towards specific capabilities and relevant general-purpose technologies. One proxy for this may be research and technologies that create insights or capabilities that are not in themselves concerning, but that could be concerning if transferred to the creation or enhancement of high consequence pathogens. I term these kinds of dual-use risks "transfer risks". These are the subject of the following analysis.

9.1 Defining transfer risks

A challenge for DURC oversight is the fact that viral engineering capabilities do not advance linearly - approaches are transferable between agents and methods are becoming increasingly general-purpose. The

horsepox synthesis debate of 2018, detailed in Chapter 1, highlights how experiments on agents with limited potential for misuse may play in advancing capabilities around the creation of pandemic agents. The main security concern around the publication of the synthesis of horsepox (Noyce, Lederman, and Evans 2018) was its relation to variola virus. Because of the relatedness between the viruses, critics argued that the same approach to assemble synthetic DNA into a functional horsepox virus may be applicable to variola virus, an approach which cost just $100,000 and only required basic laboratory equipment (Inglesby 2018; Esvelt 2018). Thus, the horsepox synthesis publication likely lowered the barrier to the acquisition of variola virus by would-be terrorists.

Research can hence feature significant dual-use potential because of its potential to translate to high-consequence agents. Risks from transferable insights have been discussed for more than a decade, including transferable enhancement methods, automation of laboratory methods, and general-purpose tools like gene editing (Steinbruner et al. 2007; National Academies of Sciences, Engineering, and Medicine 2018). Current DURC oversight is limited to research on select agents and research that could be "directly misapplied", and thus misses research that features dual-use potential because its results may be transferable to high-consequence pathogens. However, this may change in the future, as the 2023 National Science Advisory Board on Biosecurity (NSABB) recommendations include an expansion of oversight beyond research on select agents and not limiting it to research that could be "directly misapplied" (National Science Advisory Board for Biosecurity 2023). A new term for these transferable dual-use insights may be useful to assess and evaluate their implications. I will refer to these as "transfer risks", defined as follows:

> ***Transfer risk:*** *The risk of misuse of knowledge, methods, or technologies developed on (low risk) agent A which can be applied to an agent B, where B as a result poses a significant threat with broad potential consequences.*

In other words, research associated with transfer risks is not intrinsically dangerous to conduct, as it does not feature potential for accidents or misuse of physical products, but is dangerous to communicate given the dual-use potential of associated insights.[25] In contrast, research that itself involves or results in the agent of concern, such as a pandemic-capable pathogen, may be described as featuring "intrinsic risk":

> ***Intrinsic risk:*** *The risk of misuse of knowledge, methods, or technologies developed on or resulting in (high risk) agent A, where as a result agent A poses a significant threat with broad potential consequences.*

Research with intrinsic risk hence frequently is associated with physical products with misuse potential and hence is dangerous to both conduct and communicate. Current DURC oversight only captures research with intrinsic risks. One prominent example is gain-of-function work on potential pandemic pathogens like avian influenza virus. These gain-of-function studies involve experiments to select and engineer viruses to identify mutations that increase human transmissibility to inform preparedness for the spillover of these pathogens from animals (Duprex et al. 2015). These experiments involve both dangerous work with a transmissibility-enhanced virus and generate genetic information with misuse potential. Given increasing access to synthetic biology capabilities, the communication of dual-use insights from research with intrinsic risks may be sufficient to cause significant security risks (Lewis et al. 2019). Research with intrinsic risks can also feature transfer risks if resulting insights are applicable to other agents.

[25] Credit to Tessa Alexanian 2021 "East Bay Biosecurity Seminar" presentation

Transferable insights exist on a spectrum (see Figure 9.1). At one extreme lies the direct misapplication of a study with intrinsic risks, which could be defined as the most direct transfer of dual-use insights. For instance, the creation of an enhanced potential pandemic pathogen enables the creation of this pathogen in a different context, including potentially by malicious actors. At the other extreme lie the least direct transfers, such as how one study may serve as inspiration for a substantially different one. A relevant toy example is how a study looking into the spread of ideas in society could potentially inspire a virologist to develop a new method to enhance the transmissibility of a pathogen. In between these two extremes lie transfers of moderate directness, such as the transfer of experimental methods developed on one pathogen to a related pathogen. An example is how the methods for synthesising horsepox virus could be applied to smallpox virus. Studies which may enable insight transfers of moderate directness are of greatest relevance to the transfer risk concept, as associated insights may still feature significant potential for misuse but they generally are not captured by existing dual-use oversight methods. Generally, less direct transfers require more additional novel work, which in turn will involve more troubleshooting and require greater tacit knowledge - a barrier to misuse that needs to be considered when evaluating the gravity of transfer risks. I later discuss why transfer risks cannot be categorically ignored based on featuring a tacit knowledge barrier (9.5), and how defining a subset of most concerning transfer risks could make risk mitigation more actionable (9.7).

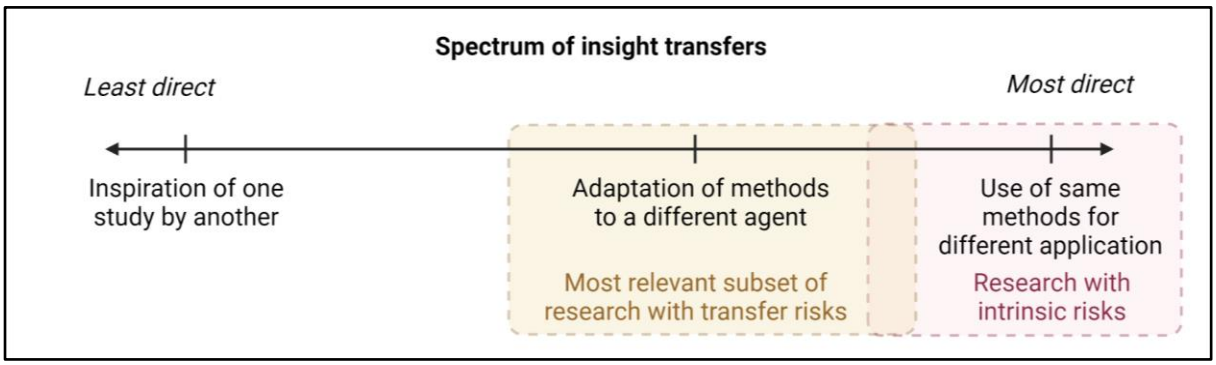


Figure 9.1: Spectrum of insight transfers and relevance to dual-use risks

Insights may be more or less directly transferred between studies and applications. The most direct transfers are the use of the same methods for a different application or the recreation of the same experiment; this is how research with intrinsic risks may pose dual-use potential. The least direct transfers are the inspiration of one study by another without any experimental methods being transferred. Moderately direct transfers could involve the adaptation of methods developed on one agent to another. Moderately direct transfers are particularly important for the present analysis of transfer risks, as they may feature a significant potential for misuse but are currently not captured by existing oversight.

Transfer risks are not new, but advances in biotechnology and an increasing focus on general-purpose methods mean that transfer risks could be growing and may become increasingly important to consider for dual-use management. In the following, I identify additional instances and examples of such transfer risks (sections 9.2 and 9.3). On the basis of these examples, I argue that transfer risks contribute significantly to advancing the capabilities to create pandemic-capable viruses (section 9.4). Subsequently, I address practical challenges to the concept of transfer risks (section 9.5). Then, I analyse the ethical implications of the distinction between transfer risks and intrinsic risks (section 9.6). I argue that the benefits and risks are not coupled for research with transfer risks and this has important implications for assessing research and risk mitigation. I finish with an outlook on how current and future dual-use management approaches may be adapted to capture the most relevant transfer risks (section 9.7).

9.2 Transfer risks from specific transferable dual-use insights

The dual-use potential of insights transferable between different agents is demonstrated by the fact that historical biological weapons programs specifically set out to find such transferable insights. For instance, the Soviet biological weapons program involved insertion and study of a range of genes in non-pathogenic vaccinia virus with the goal to later apply them to the related highly pathogenic variola virus (Gilsdorf and Zilinskas 2005). Scientists may inadvertently replicate this strategy as demonstrated by the 2001 insertion of IL-4 into mousepox. Insertion of this genetic element resulted in increased virulence and vaccine evasion of this non-variola orthopoxvirus, and these findings may be transferable to variola virus (Selgelid and Weir 2010).

My case studies show that synthetic viruses are used and engineered in ever more areas of biomedical sciences and that approaches developed on non-pathogenic viruses can be transferable to pathogenic viruses. Viral vectors for use as vaccines, gene therapy delivery methods, or oncolytic viruses are frequently engineered for immune modulation, immune evasion, and cell tropism. As I find in Chapter 5, viral surface proteins are frequently engineered to circumvent pre-existing anti-vector immunity induced by previous natural infection or vaccination (Sandbrink and Koblentz 2022). Such engineering of capsids or glycoproteins is usually virus-specific. However, when it is conducted on viral vectors that are based on attenuated versions of pathogenic viruses, such as influenza (Pizzuto et al. 2016), measles (Msaouel et al. 2012), or poliovirus (Denniston et al. 2016), engineering approaches may be directly transferable to their pathogenic cousins. This could make these pathogens able to evade vaccine-induced or naturally-acquired immunity and increase their pandemic potential (Sandbrink, Alley, et al. 2022). Transferable insights are also why the proposed development of transmissible vaccines for animal populations is highly concerning from a dual-use perspective: next to immune evasion, such research would uniquely advance potentially

transferable and generalisable capabilities for the enhancement of viral transmissibility and genetic stability (Sandbrink et al. 2021).

Furthermore, there may be certain methods for viral enhancement that are not virus-specific. Such universally applicable approaches for viral enhancement may be transferable to a wide range of agents and hence be prone to feature transfer risks. For instance, universally applicable approaches for immune evasion have been explored in the context of optimising adeno-associated virus (AAV) for gene therapy delivery (Sandbrink, Alley, et al. 2022). An example is the study by Chan *et al.,* which I described in the introduction and Chapter 6 of this thesis. This group of authors attempted to develop an approach featuring short non-coding DNA fragments that prevent detection of viruses through interfering with innate immune sensors (Chan et al. 2021). If such efforts to develop universal genetic elements for viral immune evasion achieve greater sophistication and efficacy, this will be associated with grave transfer risks.

9.3 Transfer risks from general-purpose methods

It is not just transferable and universally applicable approaches for viral enhancement that pose transfer risks, but also general-purpose methods that greatly increase the ease of viral synthesis and engineering. Fueled by novel molecular biology techniques and computational abilities, there is a growing interest to develop such general-purpose methods and platform approaches to increase the ease and speed of applying biotechnology to new challenges.

General-purpose reverse genetic platforms reduce virus-specific steps for recombinant virus synthesis. New reverse genetics methodologies are developed to increase the ease with which synthetic DNA can be turned into a functional virus. Platform approaches for reverse genetics have the ambition to be applicable across a range of viruses, for instance across multiple viral families. Thus, they may reduce specialist skills

and resources required for viral booting. For instance, Thao *et al.* developed a yeast-based platform for the assembly of RNA virus genomes and virions, which works for members of the *Coronaviridae*, *Flaviviridae*, and *Pneumoviridae* families (Thi Nhu Thao et al. 2020). The authors argue that large RNA viruses "are cumbersome to clone and manipulate" with existing approaches, and an alternative "rapid and robust reverse-genetics platform", such as their own, would be beneficial. The authors highlight the speed of their platform which can be used to create infectious viral clones in "only a week after receipt of the synthetic DNA fragments." Similarly, another group of authors advertise how their general-purpose method for the assembly of positive-strand RNA viruses works "without the need for technically demanding intermediate steps." (Amarilla et al. 2021). These reverse genetics platforms may feature transfer risks by reducing the need for virus-specific tacit knowledge that could currently provide a barrier to misuse of known viral pathogens. Moreover, accessible reverse genetics platforms could mean that many individuals will be able to synthesise high-risk pathogens as soon as they are identified - for good and ill. The development of reverse genetics platforms applicable to poxviruses might be particularly concerning from a transfer risk perspective.

Similar to general-purpose tools for viral synthesis, general-purpose tools for viral engineering feature transfer risks. I touched on some of these general-purpose viral engineering approaches in Chapter 6. Such tools include powerful methods for the directed evolution of viruses, including the use of barcoding for selecting viral mutants across multiple properties (Ogden et al. 2019; Sandbrink, Alley, et al. 2022). Barcoding describes the insertion of short sequences of DNA into viruses to identify individual viral mutants after selection. This enables parallel directed evolution of large numbers of viruses in a single experiment. Furthermore, machine learning models, such as those based on data derived from high-throughput phenotypic analyses, may be used to optimise properties like immune evasion and receptor binding (Ogden et al. 2019). Given computational approaches are frequently designed to be as flexible and

general-purpose as possible, these models exhibit transfer risks. For instance, one set of authors describes their computational approach for the engineering of adeno-associated virus as "generalizable to other proteins and engineering challenges" (Ogden et al. 2019). Computational general-purpose viral engineering methods may decrease the number of laborious experiments and the level of individual expertise required for viral engineering. They will equip an increasing number of individuals with dual-use capabilities currently limited to a handful of experts (Sandberg and Nelson 2020). While increased access to scientific experimentation brings many welcome benefits, the dual-use nature of these particular methods can pose biosecurity risks that may outweigh benefits and should be managed carefully.

9.4 Transfer risks are sufficient to reduce the crucial technical bottlenecks for the creation of pandemic-capable viruses

On the basis of the trajectory of research with transfer risks, managing these risks needs to be a crucial consideration for limiting the widespread accessibility of pandemic-capable viruses. After widespread access to mail-order DNA, there are two critical technical bottlenecks to the illicit acquisition of a pandemic-capable virus: a) the acquisition of the genetic blueprint for a pandemic-capable pathogen or rendering a pathogen pandemic-capable; and b) the difficulty of viral synthesis from synthetic DNA. As I have argued earlier in this chapter, current DURC oversight is insufficient for preventing the erosion of both protective bottlenecks as it only captures a small part of relevant dual-use capabilities. However, even if all research with intrinsic risks were subject to oversight by updated dual-use policies, this would not capture all research that is likely to enable access to pandemic viruses (Figure 9.2). On the basis of the examples discussed in sections 3 and 4, research with transfer risks will be sufficient to generate both blueprints for new pandemic-capable viruses and their synthesis, as well as disseminate relevant tacit knowledge. Thus, a governance system needs to address transfer risks to comprehensively manage dual-use

capabilities for the creation of pandemic-capable viruses. Before discussing how transfer risks might be mitigated (section 9.7), I analyse practical challenges (section 9.5) and ethical implications of transfer risks (section 9.6).

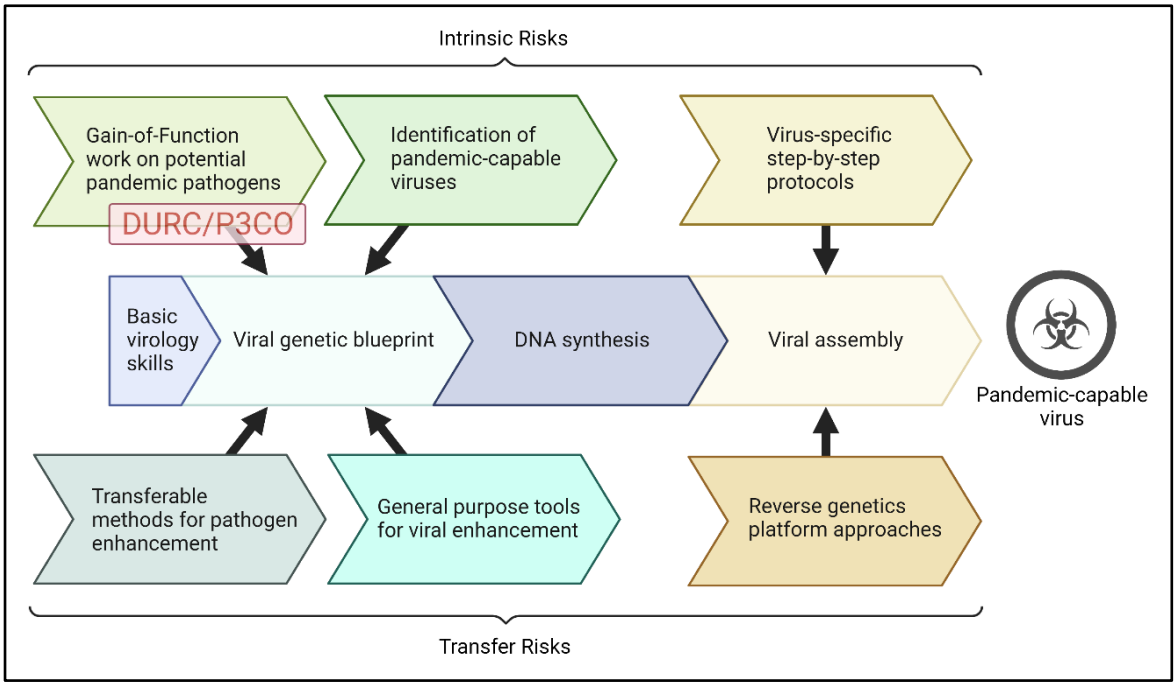


Figure 9.2: Current dual-use oversight fails to control access to pandemic viruses

Viral genetic blueprints and viral assembly are two critical bottlenecks that prevent widespread access to pandemic-capable viruses. Current US dual-use oversight DURC and P3CO policies only capture a subset of research that degrades these bottlenecks. Research with transfer risks in the form of transferable methods, general-purpose tools, and platform approaches will be sufficient for the degradation of technical barriers to the access of pandemic-capable viruses. DURC = US Dual-Use Research of Concern Policies; P3CO = Potential Pandemic Pathogen Care and Oversight Policy.

9.5 Practical challenges for the transfer risk concept

Practical challenges associated with transfer risks may limit the usefulness of this concept. First, if we cannot predict whether research generates transferable insights, the concept of transfer risks would not be useful for dual-use oversight at the research conception and funding stage. However, in many instances, transfer risks can be predicted - for instance, it was clear that the synthesis of horsepox virus would result

in transferable insights for the synthesis of variola virus. Furthermore, some of the above examples of universal and general-purpose methods are prospective and based on the prediction that certain lines of inquiry are associated with transfer risks. Many authors of relevant papers explicitly advance and advertise the general-purpose nature of their technologies, suggesting that transfer risks are not only likely to arise but also are often predictable.

Second, there may be resistance to acknowledging and considering transfer risks for fear this would result in significant barriers with significant opportunity costs of slowing relevant research. Clearly, not all research on viral vector enhancement should be subject to the highest scrutiny as this would likely mean to miss out on potent gene therapies. However, this does not mean that transfer risks of such work should not be considered as part of decisions about what research to conduct. Considering research with transfer risks requires new approaches to dual-use management, which will be touched upon in the following sections and which are the main subject of Part 4 of this thesis.

9.6 Ethical implications of transfer risks

For analysing the ethical implications of transfer risks I draw on the principles proposed by Selgelid in 2016 for guiding decision and policy-making on gain-of-function research to increase the transmissibility or virulence of pathogens (Box 2.1 in Chapter 2) (Selgelid 2016). These principles similarly capture useful ethical dimensions for evaluating broader categories of dual-use research. I have adapted a version of these principles for dual-use research more generally, which can be found in Appendix III. Here, I focus on evaluating differences in the ethical implications of intrinsic risks and transfer risks.

9.6.1 Transfer risks feature risks and benefits that are not coupled

Most of Selgelid's principles apply similarly to dual-use research with intrinsic risks and transfer risks. The balance of benefits and risks of dual-use research is generally case-specific. However, differences between intrinsic risks and transfer risks feature important implications for risk mitigation and associated ethical assessment (Table 9.1).

For research with intrinsic risks, benefits and risks are usually coupled: it is the exact same experiment, physical product, or information that results both in the majority of benefit and risk. For instance, the benefit of characterising high-risk wildlife pathogens or performing gain-of-function studies lies in the dissemination of genetic information to inform pathogen surveillance and countermeasure development - the same genetic information that enables misuse of these agents. This is because the goal of this research is threat assessment, the characterisation of future potential pandemic pathogens that might arise from nature. Most studies that involve dangerous pathogens are conducted to understand and tackle infectious disease. For this research, it is always about weighing up how much the scientific community should seek to understand the threats that are facing humanity given the potential risk of this research then enabling malicious humans to realise these threats. Therefore, the dilemma-character of the dual-use problem is particularly pronounced for the enhancement of potential pandemic pathogens, wildlife virus discovery efforts, or other dual-use research for pandemic preparedness. Risk mitigation often comes down to deciding between conducting the research or not conducting the research and potentially forsaking particular preparedness benefits (Duprex et al. 2015). Accordingly, debates on the dual-use risks of research with intrinsic risks are particularly heated and entrenched.

In contrast, research with transfer risks does not generally feature a coupling of benefits and risks. Risks are often not directly linked with the promised benefits of the research. For instance, the goal of a given

research project may be to develop a potent vector for delivering a gene therapy. The risks, however, are only linked to the particular viral vector engineering strategy that was explored to achieve this goal. Therefore, there may be alternative lines of research that have the same benefits but less transfer risks. For instance, instead of developing universal immune evasion approaches to optimise AAV gene delivery that are transferable to pathogens, AAV-specific methods of capsid engineering or pharmacological immune modulators can achieve the same benefits. In some cases, low-risk alternatives may be significantly less effective and thus will not necessarily be an attractive risk reduction strategy.

The possibility of low-risk alternatives has important implications in light of the principle of Minimisation of Risks. Selgelid bases the ethical acceptability of research on a lack of alternative less risky forms of research that would be equally beneficial (Selgelid 2016). Given the more frequent availability of alternatives, research with transfer risks may be less likely to be ethical from a Minimisation of Risk perspective. Therefore, when similarly promising projects compete for funding, one could argue that funders should support those lines of research with the least transfer risks.

Furthermore, certain general-purpose methods can be "disarmed" by limiting their harmful applications without impacting the major use case of these advances. For instance, widespread DNA synthesis capabilities constitute a general-purpose technology with great benefit for biomedical research but also harmful applications. The screening of synthesised DNA against a database of known security-relevant sequences can selectively disarm harmful applications and provide a critical layer of defence against the misuse of known pathogens (World Economic Forum and Nuclear Threat Initiative 2020). Similarly, as I have alluded to in Chapters 4 and 7, as well as other work, application programming interfaces (APIs) may be used to disarm harmful applications of computational tools for viral engineering (Smith and Sandbrink 2022). The ethical principle of Minimisation of Risks encompasses taking reasonable steps to minimise

risks of dual-use research (Selgelid 2016). Therefore, the development of general-purpose computational tools for synthetic biology should go hand in hand with the development of access systems to disarm possible harmful applications. Disarming biosecurity risks of research and technologies may be described as "biosecurity-by-design", a concept first developed for a 2016 National Defense University meeting and since expanded upon by the biosecurity community (DiEuliis and Lutes 2016; Budeski 2018). Finally, the Manageability of Risks principle states that the more manageable the risks of dual-use research are, the more ethically acceptable a study is. This may imply that certain high-risk general-purpose methods should not be developed until their risks can be managed.

9.6.2 Transfer risks and justice

Transfer risks feature unique implications for the ethical principle of Justice. According to Selgelid, justice in this context pertains to the fair sharing of benefits and risks of research; the ethical acceptability depends on the degree to which the two are aligned in their targets (Selgelid 2016). While Alta Charo argues that the benefits of gain-of-function research may be unequally shifted towards populations with stronger healthcare systems (National Research Council et al. 2015), in principle such research features alignment between the populations affected by associated benefits and risk: possible benefits in the form of preparedness against future zoonotic pandemic and risks like the accidental or deliberate release of a pandemic pathogen both affect the whole of society. For instance, the enhancement of H5N1 avian influenza virus for mammalian transmissibility generates insights with potential use for preventing global pandemics, but the associated risk of accidental or deliberate releases may also cause global pandemics. Alignment between the populations carrying benefits and risks may be characteristic of a large share of research with intrinsic risks, as such research is frequently conducted to improve preparedness against the potential pandemic pathogens it involves.

In contrast, dual-use research with transfer risks may be more likely to cause justice problems as they frequently feature a divergence between populations that carry benefits and risks. Transfer risks arise from a much broader range of research activities that aim to create benefits other than global pandemic preparedness. For instance, the use of immune evasion engineering for viral vectors to improve gene therapies would only benefit a small number of patients and the proprietors of the technology. At the same time, it is the whole of society that is affected by the risk of associated insights that might lead to the deliberate release of an engineered pathogen. It is worth noting that such a trade-off may nevertheless be regarded as just in the light of the social contract that includes care for every single individual (i.e. emphasising the significant health needs of those with rare diseases). Additionally, advances with an initially narrow set of beneficiaries may eventually lead to applications that benefit larger shares of the population.

How can we disentangle this and other ethical debates around research risks? Selgelid's principle of Good Governance - Democracy states: "Policy-making should (insofar possible) reflect the ultimate values, value weightings, and risk-taking strategies of public citizens." (Selgelid 2016). However, decision making on research risks is often done based on the values of a small group of scientists with personal or institutional conflicts of interests (Inglesby and Lipsitch 2020). Wider society and in particular those affected by possible risks need to be involved in decisions for trading off conflicting values when evaluating benefits and risks.

Table 9.1: Summary of differences between intrinsic risks and transfer risks

	Research with intrinsic risks	**Research with transfer risks**
Source of risk	Physical product and communication of insights	Communication of insights transferable to other agents
Examples	• Experiments captured by DURC, including GOF on PPPs • Laboratory characterisation of pandemic potential of wildlife pathogens	• Horsepox synthesis • Immune evasion engineering of viral vectors • Reverse genetics platforms • Computational tools for viral engineering
Coupled benefits and risks	Frequently coupled benefits and risks; same information used for benefit that creates risk for misuse.	Frequently uncoupled benefits and risks; information with misuse potential is a side effect of research.
Population targeted by risks and benefits	Usually the same population stands to benefit and carries risks, as this research is often aimed at preventing pandemics and might similarly create pandemics.	Often different populations stand to benefit and carry risks.
Risk mitigation strategies	• Decision between taking risks or forsaking benefits • Otherwise mainly communication strategies can help to manage risks	• Alternative lines of research may offer same benefits without risks • Disarming harmful applications through "biosecurity-by-design"

9.7 Mitigating transfer risks

As discussed in 9.5, research with transfer risks has many different forms, risks, and benefits. Transfer risks are representative of more diffuse advances of life sciences into directions with dual-use potential. New approaches to dual-use management that look different from existing DURC oversight are needed to manage these risks. This will be the subject of Part 4. Nevertheless, a subset of research with transfer risks may be concerning enough to be considered as part of existing dual-use oversight for high risk studies.

Export control regulations outside of the life sciences have considered transfer risks for a long time. In the nuclear non-proliferation community, it is a well-characterised risk that countries receiving support in establishing safeguarded uranium enrichment or plutonium processing facilities may use the resulting knowledge to support a nuclear weapons program (Fuhrmann 2012). Thus, the export of such facilities is generally restricted. US regulations on "deemed exports" restrict the sharing of technologies and know-how with foreign nationals in the US who may later transfer these technologies abroad (Felbinger and Reppy 2011).

Indeed, there may be movement to capture a subset of transfer risks under DURC oversight. The NSABB recommendations from March 2023 suggest removing the "directly misapplied" from the definition of DURC, as it "could limit the identification of research of concern that may pose significant threats" (National Science Advisory Board for Biosecurity 2023). If implemented in combination with NSABB's recommendation to expand review for DURC beyond select agents, then this could potentially capture experiments with particularly pronounced transfer risks relating to enhancement of viruses, including universal genetic elements for immune evasion. An important factor for effective implementation would however be the creation of heuristics to decide between what constitutes a significant threat and what does not.

To make the expansion of dual-use oversight to research with transfer risks actionable, it is important to define principles to identify transfer risks of significant concern. One approach would be to focus on transfer risks which involve potentially concerning enhancements which are readily transferable to a potential pathogen, in particular if this as a result could be turned pandemic-capable. If an experimental insight is credibly applicable to one or more pathogens, and this transfer could be conducted by anyone capable of assembling the pathogen, then the insight should be considered as concerning as if the study

had been conducted on every relevant pathogen directly. For example, if a novel method of assembling a biological agent can credibly be applied to an agent capable of inflicting severe harm and thus render it accessible to additional individuals, then it should be judged like a method with similar implications for accessibility directly developed on the agent in question. In the end, dual-use oversight should not be defined by the agent on which a given piece of research is conducted, but by its possible effect on dual-use capabilities. Such a framework would thus also capture the horsepox synthesis experiments.

9.8 Conclusions on transfer risks

In this chapter, I have discussed transfer risks as an instantiation of how more diffuse scientific activity beyond specific high-risk studies can be associated with dual-use risks. In my analysis of transfer risks, I find that studies that do not themselves have high dual-use potential could nevertheless unlock dual-use capabilities. This will only be exacerbated by artificial intelligence systems which will be trained on the whole corpus of the scientific literature and based on this input acquire relevant dual-use scientific capabilities.

Mitigating transfer risks and other more diffuse dual-use risks will require new approaches. When analysing transfer risks, I argue that they may be sometimes mitigated without forsaking research benefits. Other lines of research may achieve similar benefits with fewer risks, or complementary technologies may selectively prevent harmful applications. In the rest of this thesis, I explore governance mechanisms able to manage dual-use risks beyond high-risk studies with intrinsic risks. Such governance mechanisms will require looking beyond individual projects, to compare risks and benefits across different projects. However, existing dual-use management approaches focus on evaluating individual projects. This individual project focus, which underlies - and limits - existing governance approaches, is the subject of the next chapter.

Chapter 10: Individual project oversight and its limits

DURC review and risk mitigation strategies focus on individual projects. As discussed in Chapter 8, to comply with US DURC policies, institutions granting or receiving US federal research funding evaluate proposed research projects for select agents and experiments of concern. Researchers looking to conduct research that might constitute DURC submit their proposal to their institutional review entity (IRE), a committee of local scientific, biosafety, and sometimes legal, ethical, and regulatory experts. This review entity then evaluates whether a proposal constitutes DURC and helps to create a risk mitigation plan. However, I argue, this review process focuses on individual projects and, while it frequently involves consideration of lower-risk methods to achieve a specific goal, it only explores alternatives to a limited extent.[26] It does not involve comparisons between projects or a broader consideration of alternative beneficial projects with less risks. Thus, the US DURC policies are an example of what I term "individual project oversight".

> ***Individual project oversight:*** *Risk-benefit assessments or risk mitigation strategies that focus on an individual project, experiment, or technology.*

I differentiate individual project oversight from panoptic dual-use management, which aims to reduce dual-use risks by accounting for risks in decisions between projects and creating appropriate incentives to reduce dual-use risks. I properly introduce and expand on panoptic dual-use management in Chapter 11.

[26] Individual assessors (e.g. IRE members) may encourage the consideration of alternative or complementary projects to reduce risks. In the 2017 DURC stakeholder workshop one scientist noted that thinking about the seven experiments of concern helped identify low-risk alternatives - even for experiments not involving select agents (S. W. Evans et al. 2021).

In the following, I explore how existing oversight approaches focused on individual projects are implemented in centralised and decentralised forms (10.1). I identify that the DURC policies constitute an example of decentralised individual project oversight, while the P3CO policies attempt to establish centralised individual project oversight with moderate success (10.2). I argue that while, in the abstract, individual project oversight involves both shaping projects to reduce risks and stopping projects where risks outweigh benefits, in practice, US implementation focuses on the former (10.3). The United States could strengthen its dual-use oversight by mandating risk-benefit assessment of a broader set of research that would apply to all research independent of funding source (10.4). However, even a strengthened individual project oversight system would be intrinsically limited by not considering low-risk alternatives and the combinatorial nature of dual-use risks (10.5).

10.1 Centralised and decentralised oversight

Dual-use oversight, including individual project oversight, may be deployed in a centralised or decentralised way. Centralised oversight may be used to describe review mechanisms at an institution-overarching level, such as a central government funding body or a separate government agency. Central oversight includes scenarios where funding bodies conduct dual-use review, such as in the UK. Researchers submit their proposals, which are reviewed by institution-independent experts. In a funding body setting, individual project oversight is common, as grants are individually reviewed for risks and benefits. Centralised oversight has the advantage of being able to provide a consistent minimum bar for dual-use review. However, as there is limited room for back and forth between reviewers and researchers, centralised review processes need to be simple and uncontentious.

Oversight is decentralised when local institutions handle dual-use oversight. Such is the case in Canada, where any institution engaging in life sciences research is mandated to engage in dual-use review. Dual-use

review is frequently conducted by committees emerging from the heritage of institutional biosafety and research ethics committees. These committees have traditionally taken an individual project focus to manage safety and ethical aspects of research. Thus an individual project oversight approach is frequently prevalent in decentralised governance settings. Especially if the researcher has already received a grant, it may be difficult to consider low-risk alternatives seriously. Decentralised review has the upside of allowing increased flexibility of review processes, as local institutional reviewers may work closely together with researchers to shape projects to reduce risks. However, decentralised review processes depend on the care and skills of local reviewers and thus effectiveness of review might differ across institutions.

The downsides of decentralised institution-level review were highlighted by a 2022 controversy around Boston University experiments involving the insertion of the Omicron spike protein into the wild-type SARS-CoV-2 backbone (Chen et al. 2022).[27] When questioned about its review of associated risks and benefits, Boston University refuted risks associated with this work. Specifically, they denied that this work constituted "gain-of-function" experiments (The Brink Staff 2022) - despite the experiments featuring an enhancement of immune evasion of the wild-type SARS-CoV-2 virus. Furthermore, Boston University officials argued that this research did not have to be reported to the National Institutes of Health because it only funded relevant equipment and not the study itself. While such a lack of reporting requirement is likely true, this statement demonstrates a culture of compliance with minimal requirements, rather than proactive research oversight.

[27] This controversy is documented and analysed in an insightful manner in a Twitter thread by Marc Lipsitch (Marc Lipsitch [@mlipsitch] 2022).

Centralised and decentralised approaches may also be mixed, as is the case in Germany. The main German federal funding body DFG reviews dual-use risks of grant proposals, but also many institutions have voluntarily created local dual-use committees that work with researchers to mitigate risks.

10.2 Individual project oversight in practice in the US

Individual project oversight dominates the existing approach to US dual-use oversight. I focus on analysing dual-use management measures with an impact on what projects are funded and conducted.

10.2.1 The DURC policies

The DURC policies impose a combination of decentralised and centralised individual project oversight in the US. The 2014 iDURC policy tasks local, institutional IREs with reviewing individual projects for DURC and advising on risk mitigation measures. The DURC IRE reviews the proposed project, works with researchers to make the research not DURC, and makes an assessment of whether the research is still worthwhile even if it constitutes DURC. If a project proposal is submitted to apply for NIH funding, the institutional DURC assessment is submitted alongside the grant application.[28] Based on the 2012 DURC policy, federal funding bodies need to review all proposed intramural and extramural research for DURC.[29] To comply with this policy for extramural grants, NIH draws on the decentralised, institutional DURC review.[30]

[28] This is what relevant experts report. However, the NIH grant application instructions from 2021 do not mention dual-use or DURC at all (U.S. Department of Health and Human Services 2021).

[29] Intramural projects are those conducted by NIH scientists, extramural projects are those funded by NIH but conducted at other institutions.

[30] According to someone working at the NIH bioethics team.

Whether and how DURC review is corroborated by the funder and feeds into funding decisions is opaque. It is unclear to what extent the DURC review is factored into funding decisions;[31] this may vary based on who sits on the scientific advisory group that evaluates the merits of different proposals. After a decision to fund a proposal has been made, the outcome of the institution-level IRE DURC is declared on the NIH Notice of Award. NIH may provide support in developing an appropriate risk mitigation plan and may implement this through a term of award (National Institutes of Health 2014b).

10.2.2 The P3CO policy

The P3CO policy imposed centralised individual project oversight of ePPP research. However, its implementation is currently opaque and thus it is difficult to resolve controversy around the effectiveness of current processes. Since 2017, only three ePPP studies have been referred to the HHS P3CO committee (F. Collins 2021). Molecular biologist Richard Ebright of Rutgers University estimates that if properly implemented, tens of projects would have to be forwarded to the P3CO review committee per year.[32]

Multiple reasons may contribute to why the number of projects forwarded has been surprisingly small. First, there seems to be an incredibly high bar for a study being forwarded. Based on a yet unpublished study of ePPP research, it seems like studies are only forwarded to P3CO review if a study *clearly* meets the ePPP criteria and not if it *could* meet the ePPP criteria. This practice relegates the P3CO process from 'performing risk assessments' to 'providing guidance on managing research that someone else has deemed to feature high risks'. Second, a proposal is only forwarded to the P3CO committee if it specifically proposes ePPP research. However, a laboratory could receive NIH funding for a broader portfolio of research and equipment, and then use resulting resources (including in combination with funding from

[31] To the best of my knowledge based on reviewing official documentation and conversations with David Relman and Richard Ebright, who have both served on NIH study groups.

[32] Based on personal communication with Richard Ebright.

other sources) to conduct a study that turns out to result in pathogen enhancement.[33] This is arguably one reason why the Wuhan Institute of Virology studies to create chimeric SARS-related coronaviruses[34] were not considered by the HHS P3CO committee, and why generally there may be more ePPP studies supported by NIH funds than were sent out to the HHS P3CO committee for review (Kaiser 2021). The third factor for the low number of proposals reviewed for ePPP research may be that the P3CO policy suggests that only research is funded after a review of its merits has been performed. Therefore, NIH likely only forwards research that they want to fund - and, if only a small number of explicit ePPP projects are funded, then only a small number of studies is reviewed by the P3CO committee.[35] However, such a practice would mean that funding decisions are made before risks are properly assessed. Lastly, there is a question around how much the effectiveness of P3CO is hampered by it only constituting a departmental guideline and not a regulation that is binding.

One first crucial step to improving oversight over high-risk research is increasing transparency of the review process, which is a core goal of current discussions on the revision of the P3CO and DURC policies (Inglesby and Lipsitch 2020). The need to increase the transparency of the P3CO process features prominently in the NSABB recommendations (National Science Advisory Board for Biosecurity 2023). Furthermore, the NSABB recommendations suggest the creation of a US government office to support investigators and institutions in consistently identifying ePPP research. Given the P3CO policy is currently limited by appropriate experiments being forwarded to the review committee, this could be an

[33] If research plans change, under current practices such changes should be reported in regular grant progress reports. If a study unexpectedly results in pathogen enhancement, NIH expects for this to be reported immediately (Kaiser 2021).

[34] These studies funded by an EcoHealth Alliance grant came under public scrutiny as part of the COVID-19 origins investigations.

[35] According to personal communication with a bioethicist at NIH.

effective intervention to ensure the reliable identification of research with potential to generate high-risk results.

P3CO explicitly aspires to consider proposed projects in the context of alternatives. The policy states that there should be "no feasible, equally efficacious alternative methods to address the same question in a manner that poses less risk than does the proposed approach" (U.S. Department of Health and Human Services 2017). Indeed, for one of the three projects reviewed by the P3CO committee, P3CO-relevant activities were replaced with low-risk alternatives (F. Collins 2021). However, the current commitment to considering alternatives may be limited by its language. Nick Evans argues that asking for alternatives to be "equally efficacious" at answering the "same question" is unhelpful (N. G. Evans 2018). Any change in methodology leads to answering a slightly changed question and thus will change the benefits. For instance, if one is interested in characterising specific variants of avian influenza that are mammalian transmissible, then enhancement experiments are unique in achieving this goal. However, enhancement experiments may not be unique when a broader question is asked that aims to identify research that can help to inform policies and countermeasures against a possible avian influenza pandemic (Lipsitch 2014). Thus, Evans argues that P3CO fails to consider that alternative experiments may provide slightly different answers but nevertheless feature overall a greater expected net benefit. On an individual level, certain scientists and risk-benefit assessors may seriously consider a broader set of alternatives. For instance, in the 2017 DURC stakeholder workshop one scientist noted that thinking about the seven experiments of concern helped identify low-risk alternatives - even for experiments not involving select agents (S. W. Evans et al. 2021). However, this aspiration of comparing risks and benefits more broadly has not been formalised.

10.3 Mitigating risks with individual project oversight

Individual project oversight involves considering the risks and benefits of individual projects and engaging in relevant risk mitigation measures. After assessing risks, researchers may shape their project to reduce risks (Figure 10.1a). In some cases, risk benefit assessment may find unacceptably high risks compared to expected benefits and no plausible ways of shaping the project to achieve net benefit. Under effective individual project oversight such projects would be stopped or not conducted in the first place (Figure 10.1b).

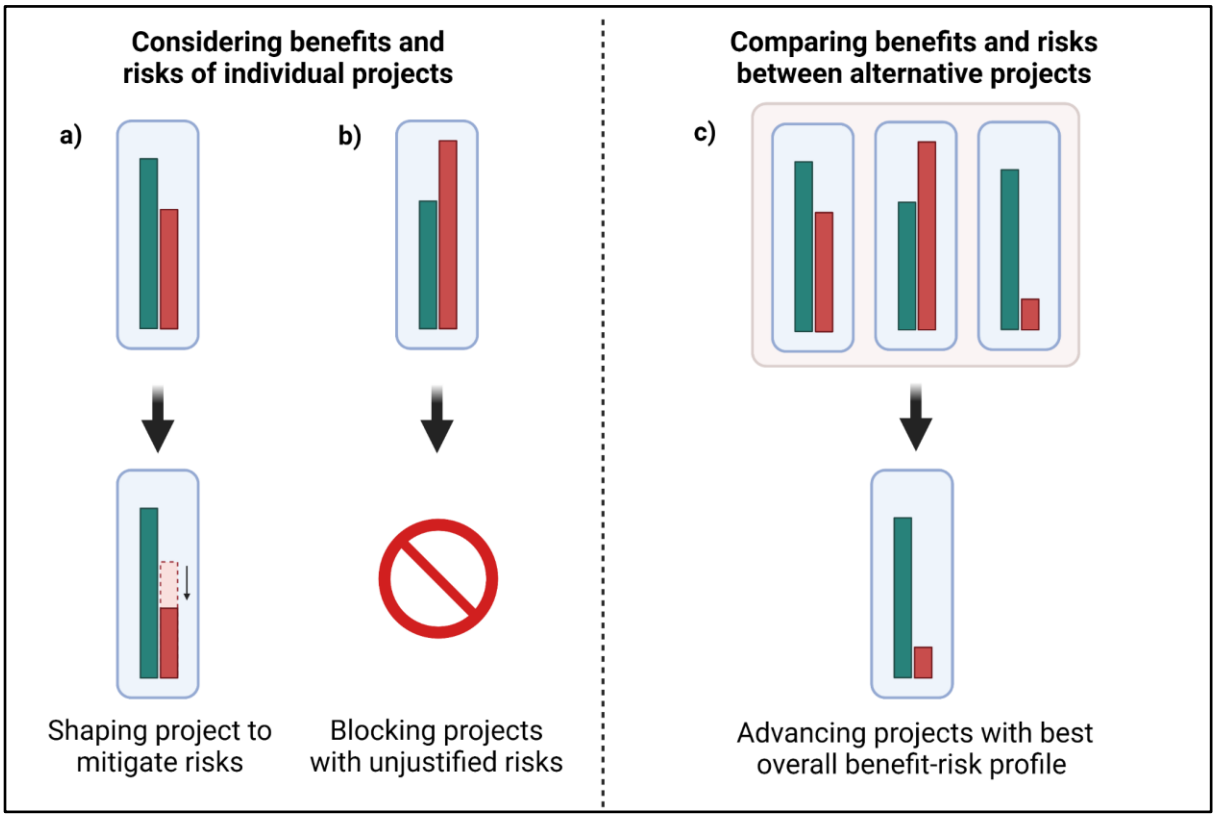


Figure 10.1: Strategies for risk-benefit assessment

Individual assessment of benefits and risks may a) identify opportunities to reduce risks through shaping a given project or b) identify risks unjustified for the level of promised benefit. c) Comparing the risk-benefit profiles of different alternative projects may enable identification and advancement of projects with overall most favourable benefit-risk profiles, which can be expected to be better than the outcome of individual risk-benefit assessment.

Within the current US individual project oversight paradigm, the shaping of individual projects is the main approach for mitigating risks. The DURC policies propose risk mitigation through modifying experiments, applying specific biosafety/biosecurity measures, evaluating medical countermeasures, and tailored communication (U.S. Department of Health and Human Services 2012). The example of a 2015 University of Chicago influenza virus research proposal demonstrates how some of these risk mitigation approaches may be applied (S. W. Evans et al. 2021, 39). To avoid the study being covered by the then-ongoing US GOF moratorium, researchers modified the experiments, applied enhanced biosafety practices, and tested medical countermeasures. The experiments were shifted to be conducted in PR8 influenza virus, an attenuated, mouse-adapted strain with reduced capacity to infect humans. Additionally, researchers committed to conducting the experiments at biosafety level 3 (BSL-3) and confirmed sensitivity of generated strains to antivirals.

The DURC stakeholder workshop in 2007 found that the most commonly employed risk mitigation strategy under DURC was tailored communication (S. W. Evans et al. 2021). Tailored communication involves presenting the research to highlight its beneficial applications while obfuscating how it may be misapplied. Tailored communication usually does not involve withholding scientific details. The NIH companion guide to the implementation of the DURC policies offers concrete advice on how to communicate dual-use research findings (National Institutes of Health 2014a). The core strategy is to focus on emphasising the value of the research and avoid drawing attention to its potential for misuse. Communication strategies may be co-developed with the NIH. If a study is ethical from a dual-use perspective, and publication of full information is justified, then tailored communication is better than 'reckless' communication, such as highlighting in a press release and in the abstract the pandemic potential of a particular novel finding. Tailored communication may also be more appropriate for publications that increase the accessibility of misuse, as this strategy might be most successful for mitigating risks from actors

with less expertise that will not realise the misuse potential of something on their own. However, tailored communication may be less suited to studies that increase peak capabilities and for which the misuse potential will be obvious. This includes the disclosure of new insights that might be particularly sensitive, such as the first set of mutations that turns a high-risk animal virus potentially human-transmissible.

Independent of the active oversight process, simply the existence of the DURC and P3CO oversight may have some incentive effect to not conduct relevant research. Because they do not want to face the burden of oversight, researchers may choose or modify their experiments so that they do not fall under the policies and their risk-benefit review. This effect has been noted for the Federal Select Agent Program, but likely also extends to DURC and P3CO (S. W. Evans et al. 2021). I will discuss the merits and downsides of these incentive effects when considering incentives more broadly in Chapter 11 and 12. However, at this stage, it is worth noting that DURC and P3CO may induce some consideration of alternative experiments before a project actually makes it to the review stage.

In certain cases, individual researchers may go beyond what is required by existing dual-use oversight. The story of the discovery of botulinum toxin H is an exemplary case of responsible communication and risk mitigation. In 2012 Stephen Arnon and Jason Barash discovered a new strain of botulinum toxin, a highly lethal nerve toxin. For fear of this new toxin being misused, Arnon and Barash excluded its genetic sequence from their initial publication (Barash and Arnon 2014). They shared the genetic sequence with a colleague to get a head start on developing an antitoxin. As public pressure mounted to publish the new sequence, researchers found that the new toxin variant was actually a hybrid between known types and could be tackled with existing antitoxins (G. Gronvall 2016, 55; Fan et al. 2016; Maslanka et al. 2016). Under existing US dual-use management, individual researchers make the decision whether to withhold security-relevant information until countermeasures are in place.

Another risk mitigation strategy available under individual project oversight is structured access to information or a technology. Security-sensitive structured access methods can either block or flag illicit access. This may be achieved through either screening for who accesses a given tool or piece of information or screening individual or combinations of queries. Examples of structured access include gene synthesis screening, which I have described in Chapter 3. Many companies voluntarily engage in gene synthesis screening based on existing US guidance. Another application of structured access are security-sensitive application programming interfaces (APIs) to mitigate the misuse of computational tools, which I described in Chapters 4 and 7. Structured access methods may be particularly important for mitigating risks from general-purpose methods or tools, which generally feature many benefits and thus are associated with high societal incentives to develop.

While existing US governance enables effective voluntary acts of risk mitigation, it features clear limitations around consistently managing risks from the most concerning subset of dual-use research. For instance, ongoing discussion about NIH's support of coronavirus characterisation and recombination experiments demonstrates the lack of conclusive governance of ePPP research (F. Collins 2021). As I argued in Chapter 9, the select agent-focused DURC policies fail to capture a lot of high-risk research. P3CO captures a broader set of work, however, as discussed in 10.2, this policy is implemented with limited effectiveness. In the following, I sketch out how US centralised individual project oversight might be strengthened to effectively manage ePPP and other high risk research.

10.4 Improving US individual project oversight

Effective centralised individual project oversight may be an important, actionable step for improving dual-use management beyond the status quo. In this section, I briefly sketch out what strengthening individual

project oversight might entail. For inspiration of how to achieve effective individual project oversight, I look to the management of other research risks and ethical challenges.

In the United States, research is regulated that involves human subjects, vertebrate animals, select agents, human embryonic stem cells, and human foetal tissue. In contrast to dual-use risks, these research areas are not just governed through guidance (see P3CO) but through federal regulations that are legally binding. The NIH review process for extramural grant application covers these risks and ethical questions in a structured and well-documented form. Grant proposals have to declare if the proposed research involves any of these materials (U.S. Department of Health and Human Services 2021). Compliance review by the NIH Center for Scientific Review ensures that all relevant items have been disclosed correctly and that paperwork, for instance required for research on select agents, is correctly presented to feed into the peer-review process (National Institute of Allergy and Infectious Diseases 2022). Dual-use research for ePPP and other high-risk research should take place with a similar level of binding formality (see Figure 10.2). For instance, ePPP oversight could be turned from soft law, characterised by self-governance and guidance-based oversight, into hard law - ensuring a solid oversight baseline for potential high-risk research in the form of risk-benefit review. Similar to how researchers should justify doing studies involving animals or humans in a grant application, researchers should also justify the use of replication-competent viruses or dangerous enhancements.

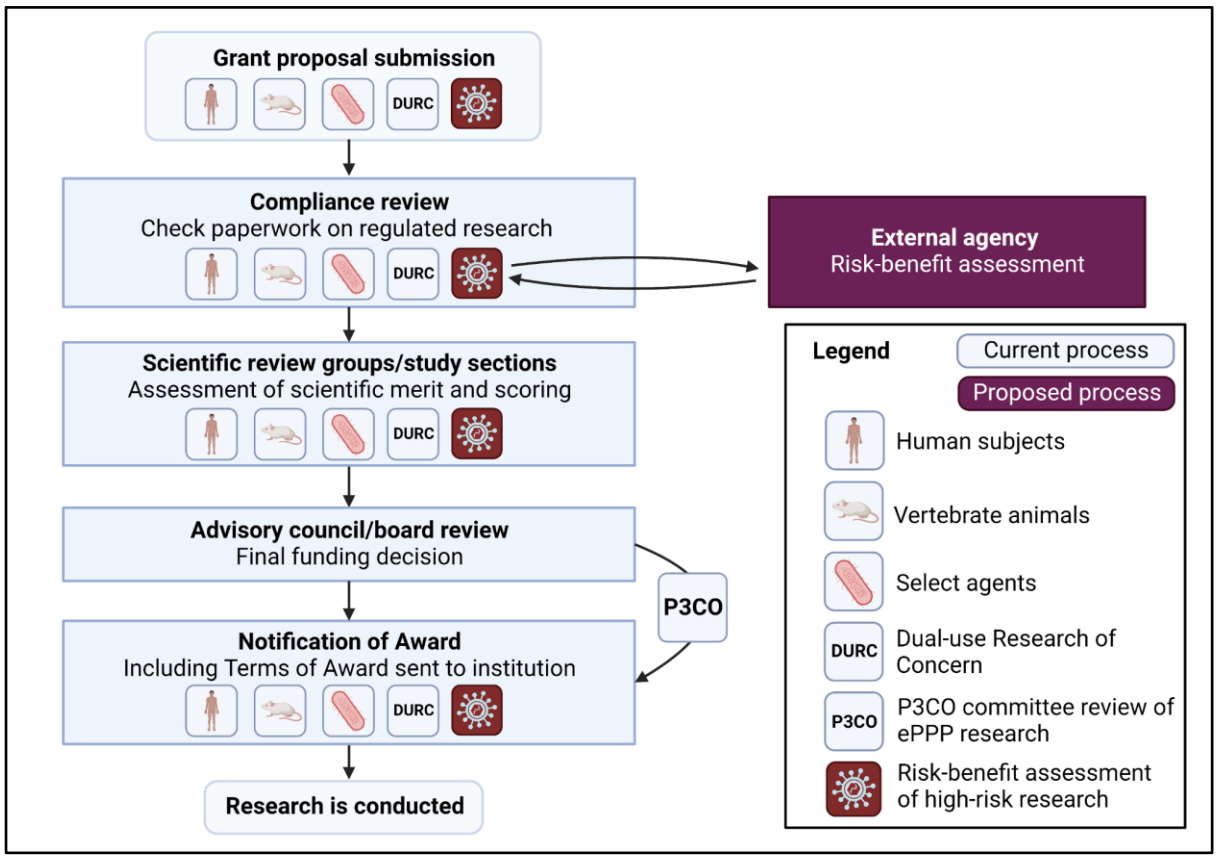


Figure 10.2: NIH extramural grant review and proposed consideration of dual-use risks

The existing NIH grant review process is pictured in blue. Before grants are evaluated for scientific merit, compliance review ensures paperwork on human subjects, vertebrate animals, and select agent research has been correctly completed. These areas of regulated research are considered by the scientific review groups and appropriate mitigation measures listed on the term of award. I propose a similar process in red for risk-benefit assessment of high-risk dual-use research. At the compliance review phase projects may be forwarded to an external review agency for risk-benefit assessment.

A committee independent from the funding body should conduct the risk-benefit assessment. The failure of effective oversight of ePPP research through an internal HHS P3CO committee has demonstrated the importance of external accountability. The US Congress, which has fiscal oversight over all US departments, is in the position to mandate such external oversight. One option could be to create a specialised, independent government body, which would review research from all funding sources (see Figure 10.2). This could include research indirectly funded government agencies - such as the Boston University experiments (see section 10.1) - and privately funded research. The central risk-benefit

assessment committee could then assess the risks of a project and whether it is allowed to proceed given what risk mitigation measures. Details of risk-benefit assessment may then also feed back into NIH to be considered by scientific review groups - this might enable comparative consideration of risks and benefits between projects, similar to existing practices for human subjects research (National Institutes of Health 2019). I explore in Chapter 13 how NIH could consider dual-use risks in funding decisions.

For effective implementation, centralised individual project oversight must have a clear scope. A clear definition of captured potential high-risk research would facilitate consistent compliance enforceable by law. Nevertheless, such a new centralised individual project oversight mechanism might go beyond the limited scope of current DURC policies. While I argue for using illustrative lists (see Chapter 8) and tiered assessment frameworks (see Chapter 13) where possible, an exhaustive list may be most actionable for centralised risk assessment required by law. An exhaustive list would simplify the process for scientists, who would simply be required to review whether their research features on a list of risky experiments. Existing ePPP and DURC research criteria may form the starting point for such a list, but it may also be expanded to address some of the limitations identified in Chapter 8. For instance, this list of most concerning activities would not have to be limited to select agents. It could also capture research to identify potential pandemic pathogens related to wildlife virus discovery. Additionally, such a list could include research beyond microbiology, such as the most concerning subset of research with transfer risks (see 9.7).[36]

[36] Indeed, NIH DURC guidelines are already more broad than legally required by the US government. This is based on personal communication with a bioethicist at NIH.

10.5 Shortcomings of individual project oversight

Strengthened individual project oversight might provide a good baseline for mitigating the highest-risk research. However, oversight focused on individual projects is intrinsically limited. Individual project oversight might not be able to achieve the best expected societal outcomes and may fail to address research with dual-use risks that do not fall into the category of highest concern, including research with transfer risks. This is because of two core shortcomings of individual project oversight: 1) the failure to consider low-risk alternatives (10.5.1), and 2) the failure to consider interactions between research projects which may shape risks (10.5.2).

10.5.1 Failure to consider alternative paths and projects

Under individual project oversight, there generally exists a high barrier to not conducting a research project. If a research project is assessed in isolation, there is a relatively high barrier to stopping it, as not doing the research may be seen as forsaking all possible benefits. In line with this, the 2017 DURC stakeholder consultation found that it is extremely rare that a project is not conducted because of its risks (S. W. Evans et al. 2021).[37]

However, in reality, the alternative to doing a potentially risky study is not to do nothing but to invest in a different research project. Other, less risky research projects may offer the same level of benefits. This may even be the case for research with intrinsic risks. If a narrow question is asked, for instance the goal to identify viruses that could spill over into humans, its achievement requires intrinsically risky findings. As discussed in the previous chapters, certain pandemic prevention research, such as pandemic virus discovery

[37] In practice, individual project oversight is quite effective at inducing consideration of alternatives as investigators often preemptively adjust projects to avoid classification as DURC/ePPP research. This is particularly the case for reducing biosafety risks in the case of research on potential pandemic pathogens.. However, such adjustments are mainly minor which means that generally the dual-use risks of a study, including the dual-use potential of the associated tacit knowledge, are not substantially modified.

or the enhancement of potential pandemic pathogens, features intrinsic risks of lab escape or informing misuse. However, for the broader question of preventing zoonotic spillovers, other research projects with less risks may offer a similar level of benefits. Instead of funding risky wildlife virus discovery research, funds could be invested into other strategies, such as detecting new zoonotic spillovers in rural communities or creating robust public health mechanisms to contain such events (Sandbrink, Ahuja, et al. 2022). When there are more promising research proposals than available funds - as is the case for preventing zoonotic spillover events that turn into pandemics - not conducting a specific study because of its risks will not meaningfully impact on the promise of the research portfolio.

Not considering low-risk alternatives may in particular hinder the governance of dual-use risks from research with transfer risks. As discussed in Chapter 9, a lot of synthetic virology research features transfer risks, which may substantially increase risks from future engineered pandemics. An example are the universal genetic elements for evading innate immune sensors developed by Chan *et al.*, discussed at different points throughout this thesis (Chan et al. 2021). Importantly, the benefits of a lot of research with transfer risks might outweigh its risks (and thus it may be judged as net positive), as it promises concrete upsides for gene or cancer therapy while producing some diffuse and indirect risks. Thus, even under appropriate individual project oversight, such research would proceed with minimal changes. However, considering alternative research avenues might identify other high-promise options with less transfer risk, options that would thus be preferable from an expected value perspective (see Figure 10.1c).

One illustrative parallel are high emissions technologies, such as combustion engine cars. Seen in isolation, a combustion engine car is net positive for society. It provides transportation, thus increasing economic output and welfare. Combustion engine cars were not banned after the discovery of their contribution to climate change, so the majority of society sees them as net positive. However, the advent of electric vehicles

has presented a more societally beneficial alternative. Electric vehicles nowadays provide similar upsides to combustion-powered cars at a lower level of environmental harm, thus featuring a greater net-benefit. Thus, governments incentivise transitioning to these vehicles and are planning to phase out combustion-powered cars over the coming decades (UK Government 2021).

Similarly, societally preferential alternatives may exist for gene therapy research for which benefits outweigh transferable dual-use risks (and thus are judged to be individually net positive). For instance, in the case of the Chan *et al.* study, the benefits of creating a potent gene therapy vector by creating universal immune modulation elements likely outweigh associated dual-use risks. However, a potent gene therapy vector may also be achieved through other paths, some associated with less risks. For instance, low-risk alternatives may be AAV-specific modifications or pharmacological adjuvant-based strategies (Sandbrink, Alley, et al. 2022). Even if low risk-alternatives do not have the very same properties, they can still feature an overall similar level of benefit.[38] Electric vehicles may have a limited range and require lengthy charging but they still feature largely similar benefits to combustion-powered cars. Additionally, they have other unique upsides such as being less noisy. Similarly, pharmacological adjuvant-based immune modulation may have different properties than genetically-encoded enhancements but nevertheless be overall expected to be roughly similarly promising. Indeed, pharmacological strategies may have unique benefits, such as the potential for adjusting dose based on symptoms. To achieve the societally best expected outcomes for synthetic virology, dual-use management needs to preferentially leverage low-risk alternatives for a broad range of dual-use projects.

[38] One limitation of this analogy is that the benefits of different lines of scientific inquiry are less certainly the same as the benefits of different forms of cars. Additionally, as part of evaluating benefits, it needs to be considered that changing research methodologies also features significant costs and risks for failure. These factors need to be considered when thinking about which avenues of research are the most promising, nevertheless, arguably, there will be cases in which two options with different levels of dual-use risks will feature roughly similar benefits.

10.5.2 Combinatorial nature of dual-use risks

Individual project oversight may also fail to consider the context that defines the risk of a given project. Projects may individually not be of high concern but in combination may enable concerning dual-use capabilities. For instance, a paper describing the genome of a novel pandemic-capable pathogen and a paper detailing step-by-step synthesis of the virus may in isolation appear less concerning. However, when evaluated in context of each other, their risks compound. Considering both projects in context of each other allows taking into account the risk created through their combination and if only publication of one of the two makes sense, evaluate which of the projects features a more favourable benefit-risk ratio.

One relevant example might be the case of the 1918 influenza virus genome published in 2005. Since 2005, as discussed in Chapter 3, multiple effective influenza virus reverse genetics protocols have been published. These protocols allow individuals with basic molecular biology skills and access to the right basic materials to reconstruct influenza viruses, which is important for improving countermeasures. However, the same protocols would now allow the reconstruction of 1918 pandemic influenza virus. It is not clear how concerning the release of 1918 influenza virus would be (Centers for Disease Control and Prevention 2019). Suppose this risk is now so high that it trumps the expected benefits of publishing the virus' genome. In that case, the potential for detailed influenza reconstruction protocols should have been considered when the risks and benefits of publishing the genome were evaluated. A similar case is the publication of the variola virus genome and the emergence of increasingly accessible orthopoxvirus reverse genetics protocols.

Individual project oversight may also fail to leverage combinations of projects that may reduce risks. For instance, as mentioned in Chapter 6, unnatural amino acids may be leveraged for intrinsic biocontainment (Rovner et al. 2015; Mandell et al. 2015). Organisms that are engineered to rely on these building blocks

which are not available in nature cannot replicate outside of a controlled setting. Thus, viral enhancement methods dependent on unnatural amino acids may feature reduced misuse risks. For instance, capsid engineering methods incorporating synthetic amino acids would mean that such modifications would not be viable for transfer to an agent designed to replicate outside a laboratory environment.

10.6 Beyond individual project oversight

In this chapter, I have argued that existing governance mechanisms, such as the US DURC and ePPP policies, focus on assessing and mitigating the risks of individual projects. I have highlighted how the United States could improve its individual project oversight. Especially promising might be binding regulations for independent, centralised risk-benefit assessment of the most concerning subset of dual-use research. However, individual project oversight features intrinsic limitations, which are especially limiting for mitigating transfer risks - risks which, as I argue in Chapter 9, may drive future dual-use capabilities. Thus, in Part IV, I explore risk mitigation approaches that go beyond the oversight of individual projects. Earlier this chapter, I already drew on the example of combustion-engine and electric cars to highlight the importance of technology governance that compares different solutions. In Chapter 11, I look closer at strategies for decarbonising the economy to inform the mitigation of dual-use risks.

PART IV: SOLUTIONS

"Regulation is the mother of invention" - Ruth Ruttenberg (Rip and Kemp 1998)

Chapter 11: Motivating panoptic dual-use management

Suppose there is a commons, a pasture shared among multiple shepherds. Each shepard faces an incentive to add animals to their herd which grazes on this commons. Adding an animal solely benefits each individual shepard; however, the costs of additional animals depleting the commons are shared among all shepherds. As each shepherd grows their herd, eventually, the commons will be overgrazed and wither. This is the "tragedy of the commons", a concept popularised by Garrett Hardin in 1968 to describe a class of population problems (Hardin 1968).

The tragedy of the commons can be found in many contemporary challenges. Antibiotic resistance, pollution, and climate change all feature properties of this tragedy. Common-pool resources may be abstract, an example being the reduction in antibiotic effectiveness which is driven by doctors with individual incentives to generously prescribe antibiotics. Characteristics of the tragedy of the commons can also be found in the dual-use problem. Examining the dual-use problem in light of the tragedy of the commons is especially important for mitigating dual-use risks beyond those with high risk for direct misapplication. In Part II and III, I highlighted how existing risk mitigation strategies fail to address risks from increasingly diffuse and transferable insights that drive powerful dual-use viral engineering capabilities. I argued that prevalent individual project oversight is important to govern the highest-risk research but is not suited to address a broader set of dual-use risks.

In the following part of this thesis, I explore risk mitigation approaches that consider dual-use risks as a negative externality of research projects, an undesirable side effect that should be considered throughout the research process - similar to how governments and companies consider carbon emissions when

deciding on what technologies to invest in, or health commissioners consider antibiotic resistance when recommending antibiotic prescribing practices. I propose exploring "panoptic dual-use management", which involves comparison of risks across research projects and alignment of stakeholder incentives with risk reduction. I define panoptic dual-use management as follows:

> ***Panoptic dual-use management:*** *Reducing dual-use risks through accounting for risks in decisions between projects and creating incentives for stakeholders to reduce dual-use risks.*

Panoptic dual-use management is inspired by the Greek word *panoptes*, which means "all-seeing".[39] At its core, panoptic dual-use management is about looking at the whole set of possible research projects and treating associated dual-use risks as negative externalities that should be considered in decisions. Through acknowledging dual-use risks as negative externalities and creating proportionate incentives to consider them, this should result in decisions for research projects that feature overall the greatest net benefit. A panoptic approach to dual-use management can be taken by a wide range of actors - from grantmakers and companies to individual researchers, as they decide between projects to advance.

Panoptic dual-use management does not prescribe how to trade off benefits and risks of different projects. Instead, it offers a methodology for making such decisions. Different from individual project oversight,

[39] One parallel which I explored but have not further substantiated is that of Jeremy Bentham's "panopticon". The best known version of Bentham's panopticon is that of a prison where cells are arranged in a circular fashion around a central guard tower. Thus, from the guard tower all prison cells can be seen, and because the guard tower does not give indication of where the guard is looking, all prisoners are constantly under the impression they are being watched. (Galič, Timan, and Koops 2017). There is some analogy to this for panoptic dual-use management: just like the guard can watch many prisoners at a time, a researcher or grantmaker should look at dual-use risks across research projects when deciding what project to advance. However, less well recapitulated is the effect of giving the prisoners the feeling of being constantly watched. A parallel could be seen in the prisoner's feeling of being watched and the creation of incentives to reduce dual-use risks across research projects for inducing a desired behaviour. However, this comparison is not only somewhat tenuous, but also controversial: it would mean that human decision makers, such as researchers or grantmakers, become the object to be surveilled. Thus, I have elected not to pursue the panopticon parallel at this stage.

panoptic dual-use management rests a risk-adjusted choice between different projects in the hands of researchers. When deciding whether to conduct a given study, individual project oversight focuses on risk-benefits assessments that generate one of three outcomes: "yes", "no", or "yes with modifications". Panoptic dual-use management adds the option of "maybe choose something else". This "maybe choose something else" hinges on comparing risks and benefits across projects and introducing proportionate incentives to account for risks.

Panoptic dual-use management thus complements individual project oversight. While individual project oversight cuts off the highest-risk research for which risks outweigh benefits, panoptic dual-use management shifts the research portfolio towards lower risk levels (see Figure 11.1). Such a research portfolio could belong to a researcher, a programme officer at a funding body, or even society as a whole. Decisions between projects that consider dual-use risks as negative externalities might result in the preferential advancement of lower-risk options or risk-reducing interventions. Such decisions may be encouraged by creating incentives for reducing dual-use risks. In theory, when dual-use risks are factored into decisions between projects, this should result in an overall shift from higher to lower risk projects.

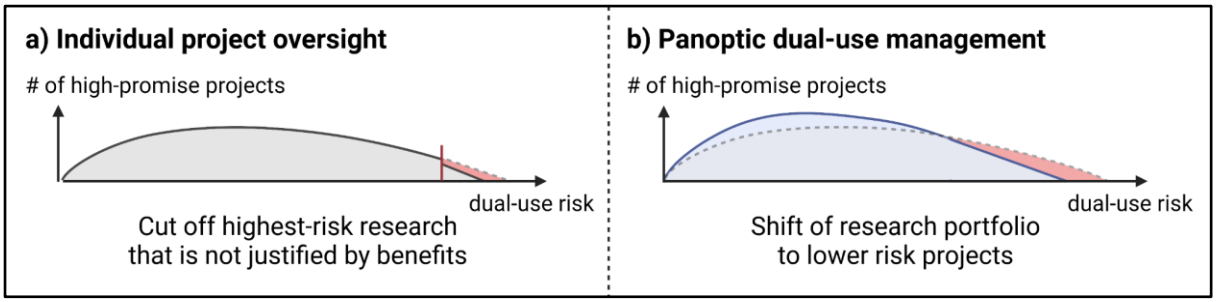


Figure 11.1: Comparison of individual project oversight and panoptic dual-use management

a) Under individual project oversight, highest-risk research is subject to oversight and is cut off if risks cannot be justified by benefits. b) Under panoptic dual-use management, interventions are used to shift the entire high-promise research portfolio to lower risk projects.

Risk assessment as part of panoptic dual-use management should also take into account that two projects together may have greater risks than the sum of their parts, and that incremental dual-use risk taken may lead to concerning capabilities that outweigh the magnitude of each individual step. This is an important aspect that should be further explored, however I here focus on the case for comparing risks between projects.

In this chapter, I discuss why a panoptic, comparative approach to dual-use risk mitigation might be appropriate and needed. I draw on lessons from climate change and the principle of differential technology development and consider possible objections. In Chapters 12 and 13, I then discuss how panoptic dual-use management might be implemented.

11.1 Parallels between climate change and dual-use risks

Climate change is a classic example of the tragedy of the commons. Climate change is driven by negative side effects of technology, carbon emissions, which individual stakeholders do not face significant economic incentives to address. Carbon emissions contribute to climate change, which has already measurable effects and features potential for increasingly devastating consequences in the future. In economic terms, carbon emissions constitute "negative externalities", defined as the consumption or production of a product causing a harmful effect to a third party - in the case of climate change, this third party being the whole earth. Tackling climate change is a coordination issue. A diverse set of stakeholders with individual incentives to keep producing carbon emissions needs to move towards reductions.

The dual-use problem features similar properties. Dual-use risks constitute negative externalities associated with life science research. For instance, for synthetic virology research that advances accessibility and capabilities, the risk is small but non-trivial that such research enables the release of a pandemic pathogen.

Kevin Esvelt argues that thousands can already today follow viral assembly protocols for influenza viruses and that these barriers continue to fall (Esvelt 2022). This will likely be exacerbated by large language models and AI lab assistants, as discussed in Chapter 7. History demonstrates the existence of individuals with the skills and motivation for starting a pandemic - an example being Saiichi Endo, the virology graduate of Kyoto University leading Aum Shinrikyo's bioweapons research in the late 1980s (Danzig et al. 2012). Thus, there is a non-trivial dual-use risk associated with viral synthesis and engineering, and these risks should be accounted for in decisions on what research to conduct.

Reducing the dual-use footprint of the research enterprise is also partly a coordination issue: even if an individual scientist would decide not to pursue a potentially risky research avenue, other scientists might still pursue the same project. Researchers and technology developers justify not taking costly measures to reduce dual-use risks by arguing that others will not do so.[40]

However, there are some important differences between climate change and dual-use risk mitigation. First, the very first disclosure of individual pieces of dual-use information may disproportionately drive the risk of misuse - I discuss this issue in section 11.1.3. Second, dual-use risks are less objectively measurable than carbon emissions and have caused little tangible harm to date. While climate change progresses slowly, the misuse of biological research may occur suddenly. These differences are crucial when considering how to structure interventions, as I discuss in Chapter 11.

Despite the challenges associated with mitigating climate change, the global community has made significant headway regarding collective action to decarbonise the economy. In the following section, I identify some learning points that might be drawn from this for the governance of dual-use risks. I

[40] From personal experience of interviewing and working with researchers and technology developers.

highlight two lessons: the importance of looking beyond individual technologies and projects (11.1.1), and creating systems that incentivise stakeholders to reduce negative externalities (11.1.2). These lessons can inform a comparative approach to dual-use management.

11.1.1 Looking beyond individual technologies

Mitigating climate change required looking beyond individual technologies. To stop climate change, society has needed to reduce carbon emissions to net zero. Modifying individual projects and technologies has not been sufficient for this goal. There is only so much you can do to increase the fuel efficiency of a combustion engine. Instead, society had to look to fundamentally less harmful technologies that fulfilled similar needs - such as advancing electric vehicles fueled by clean energy. Risk mitigation which looks beyond individual projects and technologies is the first crucial aspect of panoptic dual-use management. The following examples demonstrate that looking beyond individual projects has been crucial for tackling climate change and that similar approaches could be useful for dual-use risk mitigation.

Preferentially advancing low-harm alternatives has been at the core of progress in mitigating climate change. Clean energy sources, such as generating solar energy with photovoltaic systems, have been used to replace energy production based on fossil fuels, such as the burning of coal. To preferentially advance clean energy sources, governments have invested in research and development of relevant technologies and used policies like carbon pricing and feed-in tariffs to incentivise deployment (Green 2021; Haegel et al. 2017). For instance, Germany established feed-in tariffs in 2002 that paid owners of photovoltaic systems for generated electricity (Haegel et al. 2017). As a result, the photovoltaics market in Germany rapidly expanded, and photovoltaics companies invested in their supply chains. Such initiatives have contributed to the price of photovoltaics falling by more than two orders of magnitude over 40 years and solar energy reaching similar or even lower price levels than than fossil fuels and nuclear power (Haegel et al. 2019).

Thus, investments by Germany and others to make solar cost-effective have reduced environmental harm in the long term.

Similar interventions to advance low-risk alternatives could be imagined for reducing dual-use risks of synthetic virology. The success of mRNA vaccines for COVID-19 demonstrates the promise of non-viral gene delivery systems for therapeutic applications (Sandbrink and Koblentz 2022). As discussed in Chapter 5, advancing nucleic acid-based vaccines features less dual-use risk than engineering viral vectors. If governments removed barriers and created incentives for researchers to switch from viral vector approaches to RNA vaccines, this would reduce transferable dual-use insights and skills in the long term. A similar case could be made for the preferential advancement of non-viral delivery of gene therapy. To remove barriers to switching to non-viral gene delivery, governments could licence intellectual property for lipid nanoparticle formulation free of charge. I discuss this and other proposals for concrete interventions in Chapter 12.

Decarbonisation efforts have also looked beyond individual technologies by using interactions between technologies to reduce carbon emissions. Carbon capture technologies have been developed to complement high emissions technologies such as coal-fueled power plants. Policies for carbon pricing, a favourable regulatory environment, and financial subsidies have incentivised companies to invest in the development and deployment of carbon capture. For instance, the Norwegian offshore carbon tax resulted in the establishment of two large-scale carbon capture and storage efforts, including the world's first industrial scale effort for carbon emission abatement (Price 2014; Skalmeraas 2017).

In a similar approach, dual-use technologies can also be complemented with risk-reducing technologies. One example are structured access technologies such as gene synthesis screening, which I have discussed in

depth in Chapter 3. Another example of a risk-reducing complementary technology might be non-canonical amino acids. In Chapter 6, I introduced the idea of using amino acids that are not normally found in the human body to reduce the dual-use potential of viral engineering approaches for gene therapy. Viral enhancement methods incorporating non-canonical amino acid approaches cannot be used to enhance viruses designed to replicate outside of a controlled environment.

11.1.2 Alignment of incentives

The second crucial part of panoptic dual-use management is using and aligning incentives to account for harms associated with different projects and technologies. Economic theory predicts that when decision-makers do not carry risks, more risk is taken than is societally beneficial (Arrow 1970). In the tragedy of the commons, incentives among shepherds are misaligned: individual shepherds face incentives to add animals to their herd, despite such actions being overall harmful to the whole community of shepherds. Similarly, misaligned incentives also exist in climate change and the dual-use problem.

Companies and consumers generally do not want to pay the costs of reducing carbon emissions if no one else pays a similar cost. A consumer survey from 2020 revealed financial costs and other negative short-term impacts as the biggest hurdle for decarbonisation (Ofgem and Revealing Reality 2020). Consumers worried that their efforts would go to waste if others would not engage in similar activities. Some surveyed individuals said they do not feel responsible for reaching net zero and assigned responsibility to governments and energy companies.

Just like consumers are restrained by economical barriers from individually reducing carbon emissions, researchers are restrained by career incentives from reducing dual-use risks. Researchers are subject to

incentives for career advancement, such as receiving grants and publishing high-profile papers.[41] If a researcher would consider dual-use risks and their mitigation, this might disadvantage them in the short-term. Considering dual-use risks might complicate their work and cost them time. Furthermore, if a given researcher would decide not to conduct a given experiment, then others that are less conscientious might and thus receive the laurels. In fact, researchers may face incentives to publish more controversial research. If a study is more controversial, it may gain greater publicity and high impact journals might be more likely to publish it. For instance, the Boston University SARS-CoV-2 spike recombination experiments discussed in Chapter 10 ended up being published in *Nature,* after their provocatively written preprint sparked discussion about risk management (Chen et al. 2023; Salzberg 2022; The Brink Staff 2022)..

Both mitigating climate change and dual-use risks feature misaligned incentives and thus result in a coordination problem. A potent intervention would be to align incentives faced by individual decision-makers with the true long-term societal costs and benefits of different options. In economic terms, aligning incentives in this way is called "internalising externalities" (Owen 2006). As discussed in the previous section, policies that leverage incentives have been used to advance clean energy sources and carbon capture. Such policies include feed-in tariffs to incentivise the adoption of photovoltaics and carbon pricing to incentivise switching to low-emissions technologies or advancing carbon capture storage. Similar policies could be envisioned to incentivise researchers to preferentially pursue low-risk projects and engage in risk-reducing practices. I discuss how such panoptic dual-use management interventions might look in Chapter 12.

One aspect complicates fixing incentives for climate change and dual-use from an ethical perspective: there may be a temporal lag between costs and risk reductions, and the identities of individuals benefited and

[41] Based on personal conversations with a number of practising scientists.

burdened may differ. Concretely, interventions to align incentives place costs on individuals now for the benefits of potentially different individuals in the future. This is especially a problem in climate change, as time horizons are relatively long; while early effects of climate change are already noticeable now, an assessment of catastrophic climate change focused on 2060-2080 (Kemp et al. 2022). Individuals older in their 40s or 50s in the 2020s will likely not be around to bear the brunt of climate change. Temporal lag and a divergence of those affected might be less of a problem for dual-use risks. Assuming that catastrophic biotechnology misuse may happen in the next decades,[42] even researchers and other stakeholders at the peak or end of their careers might still be affected.

Beyond temporal lag, there are also other ethical questions to consider when weighing costs and benefits of interventions to manage dual-use risks. Indeed, my discussion of transfer risks and justice in Chapter 9 (9.6.2) highlights that who is benefited by a given research project may also differ from who is affected by risks. For instance, research projects on developing efficient viral vectors for gene therapy may only medically benefit a small subset of the population, while potentially contributing to enabling a deliberately-caused pandemic affecting the majority of society. It is important to distinguish true societal costs of developing one technology over another from those costs for individuals that arise as part of zero-sum competition such as career advancement. Assuming similar moral weight of individuals across geography and short timeframes of decades, shaping incentives of individuals that make decisions with an effect on others can generally be justified - even if whether an intervention is proportional depends on its details.

[42] The timeframe with the greatest risk for misuse may be 10-30 years from now when abilities to create pandemic viruses have become even more accessible and powerful, however pandemic response technologies have not yet had time to catch up.

11.1.3 Considering information hazards and the Unilateralist's curse

A crucial way in which the negative externalities from dual-use research differ from that of high-emission technology are information hazards and associated disproportionate first-disclosure risks. The first instance of a given dual-use experiment or finding may be argued to create the majority of misuse risk. This is especially the case if the majority of the risk is not associated with the physical products or skills that a project produces, but with dual-use information which poses information hazards (Lewis et al. 2019). For instance, work that identifies a pandemic-capable virus features some risk that physical pathogen samples will be misused, but poses also the much greater risk that one of thousands of individuals might recreate the pathogen from scratch if its genome is disclosed. Dual-use information poses risk depending on its novelty. If the genome of the pandemic-capable pathogen is the first of its kind that is published, this hands a new dangerous agent into the hands of the world. In contrast, follow-on work that identifies functionally similar variants of a pathogen already known would feature less dual-use risks. Thus, often the first disclosure of a piece of information is particularly hazardous while subsequent disclosures feature less additional risk.[43]

This disproportionate first-disclosure risk is made even more challenging due to the Unilateralist's curse: if a number of actors individually decide on whether to conduct a piece of research or not, variance in judgments mean that the research is conducted more frequently than socially desirable (Bostrom, Douglas, and Sandberg 2016). For instance, if a number of researchers consider the benefits and risks of publishing the genome of a newly identified pathogen, judgements will differ. Even if most researchers judge that the risks of disclosure outweigh its benefits, because of imperfect information accessible to any single individual, a single researcher may mistakenly judge benefits higher than risks and publish the genome in

[43] Raising additional attention to information hazards is also concerning, Nick Bostrom has termed this "attention hazards" (Bostrom 2012).

question. Incentive-based dual-use management will suffer from the Unilateralist's curse. Even after incentives have been adjusted, researchers will make independent assessments of how risks and benefits trade off. Thus, researchers might come to divergent conclusions and projects might be conducted even if they are unfavourable compared to alternatives after consideration of risks.

However, this does not mean that there is no point to adopting panoptic dual-use management and adjusting incentives for mitigating dual-use risks. Rather, the Unilateralist's curse highlights the importance of additional individual project oversight. By definition, first disclosure of a piece of dual-use research is particularly concerning for the highest-risk research - research with insights that can be directly and independently misused, such as the publication of the blueprint for a pandemic-capable virus. For such highest-risk research, individual project oversight can reign in the Unilateralist's curse by providing a mechanism for judgements that are not unilateral but integrate a range of perspectives. Meanwhile, a larger fraction of the dual-use risks of a more diffuse set of research might be related to the generation of tacit knowledge and the generation advancement of a research area, which are less affected by first-disclosure risks. Thus, the Unilateralist's curse may be less limiting when targeting a wider range of dual-use research, as is the main aim of panoptic dual-use management.

Taking a comparative approach and adjusting incentives can add value on top of individual project oversight and despite being potentially limited by the unique properties of information hazards and the Unilateralist's curse. Adjusted incentives can help to ensure that stakeholders consider risks during a wider set of decisions. Incentive-based panoptic dual-use management is particularly important for capturing a larger set of diffuse and incremental dual-use advances, including capturing dual-use risks that do not suffer from the Unilateralist's curse, such as spreading dual-use skills for viral engineering. I will discuss

more how to integrate the different governance approaches and ensure the greatest global coverage of dual-use risks in Chapters 13 and 14.

11.2 Differential technology development and the role of defensive technologies[44]

Next to climate change, the other inspiration for panoptic dual-use management is the principle of differential technology development. Differential technology development highlights that the relative timing and sequence in which technologies are developed have important implications for risks. Technology risks are shaped by the existence of other technologies. For instance, cars are made substantially safer by seat belts and airbags. When cars were first adopted, seat belts and airbags did not exist and thus, accident-related fatalities were greater. If seatbelts or airbags had been developed alongside cars, they could have been integrated sooner and thus have reduced accident-related fatalities.

The principle of differential technology development was first articulated by Nick Bostrom (Bostrom 2014). Bostrom introduces "differential technological development" to highlight that advanced artificial intelligence might reduce risks from other technologies. Thus, despite being associated with risks itself, artificial intelligence should potentially be developed sooner than later. With colleagues, I have fleshed out this principle and put it in the context of the responsible innovation literature (Sandbrink, Hobbs, et al. 2022). There are challenges around how to implement differential technology development without falling into technocratic pit holes.[45] However, its framing of technology interactions nevertheless provides important lessons for dual-use management.

---

[44] This section features content from a co-authored preprint: Sandbrink, Jonas B., Hamish Hobbs, Jacob Swett, Allan Dafoe, and Anders Sandberg. 2022. "Differential Technology Development: A Responsible Innovation Principle for Navigating Technology Risks." SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=4213670.

[45] For example, one critical challenge is how much to value different risks and benefits of various technologies. Different societal groups might have different perspectives on this. Thus, operationalising differential technology development requires careful consideration of how to consider divergent values across society.

Differential technology development highlights how risks from technologies depend on their context. There are different forms in which technologies can interact to reduce risks (see Figure 11.2). The examples discussed above for climate change illustrate two classes of such interactions. First, integrated carbon capture and gene synthesis are examples of safety technologies, technologies that modify other technologies to reduce associated harms (see Figure 11.2a). If a safety technology is developed and adopted earlier, this can reduce a window of vulnerability. Second, the replacement of fossil fuels with clean energy sources highlights how low-risk solutions can reduce the need to develop and deploy higher-risk technologies for similar applications. Similarly, advancing low-risk mRNA vaccine platforms reduces the need for viral vector approaches (see Figure 11.2c). If low-risk solutions are advanced earlier, this reduces the need to explore or use higher-risk alternatives.

One last set of technology interactions has not been substantially leveraged for reducing climate change. These are defensive technologies which reduce harms from other technologies without directly interacting with them. For climate change, an example would be stratospheric sulphur injections for geoengineering - a strategy largely abandoned for its downside risks (Stilgoe, Owen, and Macnaghten 2013). However, for preventing pandemics, defensive technologies like vaccines and diagnostics are very critical and popular interventions. Advancing defensive technologies earlier relative to dual-use capabilities shortens a window of vulnerability (Figure 11.2b). In the last chapter, I discussed how Stephen Arnon delayed the publication of Botulinum Toxin H to allow testing of an antiserum. This case exemplifies how the relative timing of a dual-use capability and a defensive technology can be used to shorten a window of vulnerability.

Risk-reducing interactions between dual-use and defensive technologies highlight an important lesson for dual-use risk assessment and mitigation: the level of dual-use risk depends on the point of assessment; a

capability that features high dual-use risk now, might in the future feature less risks because of advances in defensive technologies. If society learns how to robustly contain pandemics faster than creating pandemic-capable pathogens, this reduces misuse risks of future virus engineering research. Thus, deprioritising a project because of its risks does not necessitate that this project should never be conducted - risks may well be lower in the future. One effect of this is also that if a dual-use management strategy is only transiently effective, it can still be a worthwhile strategy. An example is gene synthesis screening. Once computational tools for the bottom-up design of pandemic-capable pathogens are available, existing gene synthesis screening approaches based on known pathogens will fail. Nevertheless, current approaches for gene synthesis screening are important tools to manage misuse risks over the near future.

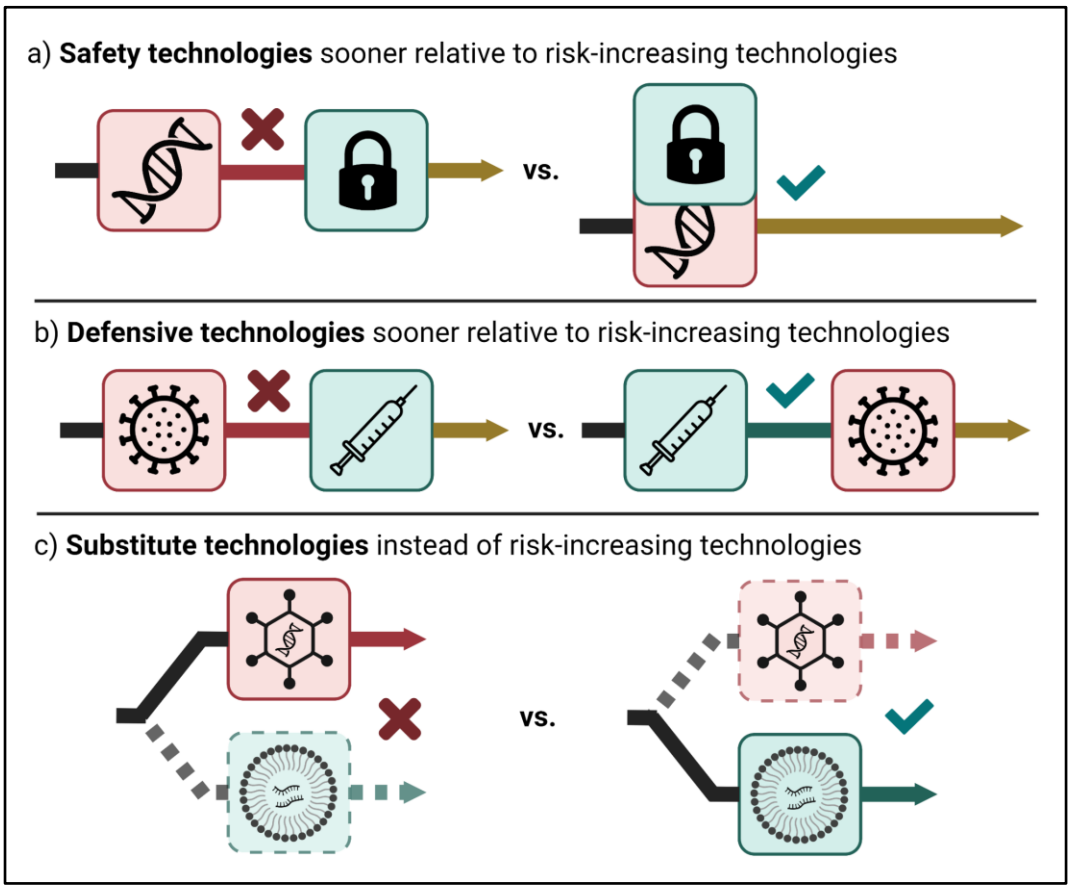


Figure 11.2: Risk-reducing technology interactions in synthetic virology

a) Developing safety technologies with or soon after risk-increasing technologies can reduce risks, e.g. gene synthesis screening may prevent the illicit synthesis of harmful pathogens. b) Developing defensive technologies before or soon after risk-increasing technologies reduces a window of vulnerability; e.g. society should first develop ways to prevent

a pandemic from a particular virus, for instance through vaccinating the population before disseminating its blueprint for its synthesis for biotechnology applications. c) Low-risk technologies can replace higher risk solutions or make their development unnecessary; e.g. RNA vaccines are a low-risk alternative to viral vector vaccines. Adapted from Figure 1 in (Sandbrink, Hobbs, et al. 2022).

11.3 Possible objections to panoptic dual-use management

In this chapter, I introduced the concept of panoptic dual-use management, a dual-use risk management strategy that focuses on establishing incentives for stakeholders to consider dual-use risks in decisions between projects. I argued that this strategy would complement existing individual project oversight by enabling dual-use management across a broader set of research and a move towards a research portfolio that more accurately tracks the true (dual-use risk-adjusted) expected value of projects.

At this stage, it may be worth considering possible objections to the need for panoptic dual-use management. I already discussed challenges from information hazards and the Unilateralist's curse in section 11.1.3, concluding that incentives to consider dual-use risks in decisions can nevertheless be worthwhile if coupled with individual project oversight for the highest risk research. I will discuss the implementation of incentives and associated challenges in Chapter 12. In the following, I address different higher-level objections.

11.3.1 Are existing approaches for dual-use risk mitigation sufficient?

One objection to panoptic dual-use management could be that existing dual-use oversight approaches already have incentive effects and they leave scope for capturing emerging dual-use risks. Arguably, individual project oversight could be expanded to the emerging research areas featuring significant dual-use risks, including projects with concerning transfer risks. Indeed, the new NSABB recommendations may be taking a step in this direction (National Science Advisory Board for Biosecurity 2023).

Additionally, individual project oversight is already creating incentives to preferentially pursue projects that would not be subject to the oversight, because of associated paperwork and other administrative burdens. Furthermore, institutional review entities already often consider low-risk alternatives when reviewing proposals for dual-use research of concern (S. W. Evans et al. 2021).

The counter to this objection is that panoptic dual-use management does not challenge or ignore the impacts of existing dual-use oversight. It does not necessitate departing from existing mechanisms. Rather, it makes a mindset explicit that considers dual-use risks as negative externalities and accounts for this by correcting incentives. Thus, panoptic dual-use management can make the use of incentives more explicit and deliberate. For instance, it could create a structured mechanism to consider, shape, and review the incentives that oversight practices are already having on what research projects are pursued - including ensuring that no research areas are inadvertently neglected.

Additionally, panoptic dual-use management offers the potential to actually consider and manage the risks from a larger set of projects. As discussed in Chapter 10, existing individual project oversight involves evaluating whether a project risks that may outweigh its benefits and adapting the project so that risks are reduced. In contrast, panoptic dual-use management can also manage risks from projects where benefits outweigh risks when considered in isolation. Such risks should still be considered because there may be alternative projects with less risks and greater overall net benefit, because delaying conducting or publishing research may reduce risks (see Botulinum H example discussed in 10.3), or because in sum with other parallel projects risks may become excessive.

11.3.2 Incentive effects could prevent important work from happening

A second objection to panoptic dual-use management is that associated incentive effects to reduce dual-use risks could prevent important work from happening. While generally creating incentives to reduce dual-use risks seems sensible, this can easily cause harm if as a result certain important work is neglected. For instance, there has been a feeling that the US DURC policies have reduced work on specific pathogens that would be critical for improving medical treatment.[46]

At the core of panoptic dual-use management lies the ambition that *after* accounting for dual-use risks the research with the overall greatest *net benefit* should be conducted. Thus, if panoptic implemented well, no important work should be unjustifiably discounted. Thus, this objection does not question panoptic dual-use management *per se*, but rather questions whether it could be implemented in a way that actually has the desired effects.

How to implement panoptic dual-use management is the subject of Chapters 12 and 13, including challenges of implementation such as ensuring proportionate incentive effects, coordination across jurisdictions, and inadvertently creating adverse reactions. As the proposed different ways to shape incentives are trialled, it is important to monitor their effects and finetune implementations based on this. Arguably, making dual-use management explicitly about incentives will help detect undesired negative effects, such as those currently associated with the DURC policies.

One related implementation challenge will be how to consider differences in opinions and values across society. For instance, this will be important when making judgements about the extent of different risks and the value of mitigating them. As part of designing and administering panoptic dual-use management

[46] Based on personal communication with a senior microbiologist and infectious disease physician.

and related incentive mechanisms, it will be important to engage diverse stakeholders including the general public. Such broad stakeholder consultation would inform the general principles for how to value and trade off different risks and benefits. On the basis of these principles, a combination of experts and administrators can then make decisions on the details of implementation. Notably, this is not different from how decisions on the benefits of research projects are made as part of funding processes at governmental research funders. As I describe in more detail in Chapter 13, the National Institutes of Health features scientific review groups consisting of different topic-level experts which score different research proposals.

11.3.3 Other improvements to dual-use oversight take greater priority

The last objection that I discuss here is that while panoptic dual-use management can help tackle more dual-use risks, other dual-use risk mitigation advances are of greater urgency and priority. As the 2021 Global Health Security Index finds, the majority of countries do not even have the most basic dual-use oversight of high-risk research (Nuclear Threat Initiative and Johns Hopkins Center for Health Security 2021). Furthermore, specific governance of crucial enabling technologies like DNA synthesis and artificial intelligence tools may be the most important lever for managing misuse risks.

Even if these other goals for reducing dual-use risks are more urgent and important, which is plausible, exploring panoptic dual-use management may still make sense. First, even if the immediate policy priority is enhancing oversight of the highest-risk research, advancing ideas for panoptic dual-use management makes sense to enable their consideration, refinement, and possible eventual implementation. Indeed, the need for panoptic dual-use management may grow. One mechanism might be that wildlife virus discovery efforts and advances in artificial intelligence make it easy to identify and design pathogens capable of causing a pandemic. This would mean that basic synthetic virology skills become the crucial factor for

enabling the deliberate release of a pandemic pathogen; thus, misuse risk may almost be directly proportional to the number of individuals trained on a broad set of projects. Another mechanism may be that artificial intelligence systems become very good at combining the results of all conducted experiments into broadly accessible scientific capabilities, so shaping the direction of a broader set of research will be crucial to shape the dual-use potential of the science capabilities of these systems. Second and importantly, panoptic dual-use management may make the governance of high-risk research and general-purpose enabling technologies easier. For high risk research, creating incentives and enabling comparisons across projects may be effective at encouraging alternative experiments - and may be received more favourably than the decisions of risk-evaluating bodies on whether an experiment should be conducted or not. Incentive systems may also be useful to realise the governance of general purpose technologies, as I discuss for DNA synthesis screening as part of the next chapter.

Chapter 12: Panoptic dual-use management through incentive systems

"Prohibition is easy to legislate; but how do we legislate temperance?" - Hardin 1968

The core tool for realising panoptic dual-use management is the use of incentives. Incentive systems, including disincentives, could nudge stakeholders to consider dual-use risks when deciding what research to conduct or fund. Similar incentive systems are already in use for reducing carbon emissions. Incentives or disincentives through taxation or social pressure can sway companies to reduce environmentally harmful emissions. Currently, researchers make decisions about their research without direct incentives to consider the low-probability, high-consequence societal costs of dual-use research. Incentives could ensure implicit consideration of dual-use risks in decision processes. Some existing regulations, such as requirements for select agent research, already have incentive effects (S. W. Evans et al. 2021). However, in these cases, incentive effects are frequently unintentional side effects and not optimised for inducing desired behaviour changes. For instance, the US select agent regulations have the primary goal of ensuring the physical security of public-health relevant pathogens and toxins through requiring registration and security screening of laboratories and researchers engaging in relevant work. Although this is not their purpose, associated paperwork and wait times disincentivise new laboratories or researchers to engage in research on select agents.

Using incentive systems to shape scientific advances fits well into Kathleen Vogel's biosocial frame, which I introduced in Chapter 2. Vogel argues that the trajectory of biotechnology is not fixed or technically deterministic. Thus, interventions on what research is conducted can shape the trajectories of science and technology in more or less positive ways (Vogel 2012). Incentive systems may nudge humans to deploy their agency towards shaping science to reduce risks.

Incentives may take financial or non-financial forms (Lee 2015). Financial incentives usually relate to economic loss or gain and may include taxes and tax benefits. Financial incentives can internalise externalities and translate the true societal cost or gain of different options into a decision process (Owen 2006). Non-financial incentives include social pressure and reputational gain or loss among peers and the general public. Financial and non-financial incentives may work best in different contexts. Financial incentives have worked best to achieve carbon reduction targets in companies (Ott and Endrikat 2022) and have worked to reduce antibiotic prescribing (Bou-Antoun et al. 2018), while non-financial incentives have excelled at encouraging healthcare staff to provide better patient care (Lee 2015). Farquhar *et al.* have previously proposed using financial incentives to price in safety and security risks of dangerous pathogen research (Farquhar, Cotton-Barratt, and Snyder-Beattie 2017). Where relevant, I draw on their proposals of mandatory liability insurance and centralised risk-assessment-based risk pricing.

Implementing incentive systems for risk mitigation is made easier by multiple factors. First, such an approach only requires an estimation of risks, which Gryphon Scientific found easier than the estimation of benefits (Gryphon Scientific 2016). Second, decisions on what projects to conduct remain with researchers. Thus, incentives for risk reduction do not impose more on academic freedom than existing incentives, such as career incentives to publish in high-impact journals. Third, governments control a lot of research funding, and thus have significant leverage to reduce dual-use risks - in this aspect, dual-use risk mitigation is more actionable than the reduction of carbon emissions. I discuss in detail how government funding bodies could create incentives to reduce risks in Chapter 13. Fourth, policymakers could develop incentive systems with broader public input on valuing risks, replacing individualistic judgements on the costing of dual-use risks. To make societally beneficial decisions, researchers currently have to become risk assessment experts (Palmer, Fukuyama, and Relman 2015). Building incentive systems would alleviate this challenge and could hand difficult questions on the assessment and costs of risks to diverse societal voices

and relevant experts of science, biosecurity, and bioethics. Lastly, there is encouraging precedent from other factors relating to science where incentives have led to significant improvements. Funding bodies requiring researchers to consider the 3Rs when proposing animal experiments - Replacement, Reduction, and Refinement - has significantly spread awareness of the need to minimise animal suffering in research (Hubrecht and Carter 2019). Incentives have also been used to advance Open Science and drive science towards greater transparency, examples of which I use as inspiration in later sections of this chapter.

12.1 Challenges for implementing incentives to reduce dual-use risks

There are significant practical challenges for implementing panoptic dual-use management and establishing incentive systems for reducing dual-use risks. The first challenge is specific to incentives for dual-use risk mitigation, the others are general challenges of any incentive system.

12.1.1 Predicting and measuring dual-use risks

When using incentive systems to internalise dual-use risks into stakeholder decisions, it is important to create incentives that are proportional to the magnitude of risks - so as not to overshoot or undershoot. However, in contrast to carbon emissions which can be priced by the metric ton, dual-use risks are difficult to measure and quantify. Even among biosecurity experts, evaluations of risks differ widely.[47] Given misuse events are rare but potentially catastrophic, it is hard to establish feedback loops on what actions are proportional.

The difficulty in measuring dual-use risks limits how readily lessons for creating incentives can be applied from other areas. However, such parallels may still serve as inspiration, and indeed some areas have created

[47] Forthcoming paper by Tessa Alexanian and Daniel Greene on iGEM project risk assessment.

solutions that still might be suitable. There is precedent for incentives aimed at reducing adverse events with small probabilities and large consequences. For instance, in the US, nuclear power plants are required to take out liability insurance for accidents and terrorist attacks. In this model, calculating difficult-to-quantify risks is handed to insurance companies. Another approach for addressing difficult-to-quantify dual-use risks may be comparison of relative rather than absolute risks. Relative risk comparisons may be very actionable for research projects within a given topic area, as I discuss in more detail in Chapter 13.

Even if it is difficult to ensure incentives are proportional to risks of individual experiments, incentive systems can still be useful. In this case, it is important to judge whether overall efforts and incentive systems are proportional to their goal of addressing the security risks of capabilities to start a pandemic. Currently, given society's vulnerability to pandemics and very limited efforts to reduce misuse risks, improving dual-use oversight is likely neglected.

12.1.2 Competition and coordination

Competition generally causes technology to move along more deterministic paths (Dafoe 2015). Competitive dynamics constrain the decision space of individual actors, a dynamic that may apply to competition between countries to win a market or competition between researchers to get a high-profile publication. A race mindset may prevent consideration of risk-reduction strategies that involve slowing or increasing the cost of risky research. The September 2022 US executive order on strengthening the bioeconomy may be an example of this (The White House 2022). In cases where race dynamic exists, using incentives for low-risk alternatives features more promise than disincentives proportional to dual-use risks.

Competition is partly a coordination challenge. If one government puts disincentives for high-risk research in place and other countries do not, research might move elsewhere. I discuss how coordination may be improved - and why it might not be necessary - in Chapter 14.

12.1.3 Adverse incentives

Any incentive system needs to avoid creating adverse incentives. If declaring dual-use risks is associated with a penalty, researchers may be less inclined to do so. If dual-use research is associated with additional costs or regulations, researchers might be incentivised to underestimate risks. Therefore, incentive systems need to be simple to implement and designed with unambiguous and effective rules. Additionally, some research suggests that "performance pay" may decrease intrinsic motivation to mitigate risks. However, this evidence is very context-dependent, and these negative effects are avoidable with careful implementation (Kunz and Pfaff 2002). Based on theoretical and experimental evidence from business environments, Kunz and Pfaff argue that rewards generally only have detrimental effects when there is high initial interest in performing a task (generally not the case for reducing dual-use risks), there is a lack of surveillance for continued performance (easily avoidable when creating systems), and the reward is for actions that are generally not compensated (limited precedent for dual-use risks given very few researchers actually actively engage in it, but could be a concern for a subset of researchers). Kunz and Pfaff also find that incentive systems ideally allow for performance improvement, thus incentives for dual-use risks mitigation might consider how exceptional performance could be rewarded. Lastly, Stuart Buck describes a risk of "performative reproducibility" for incentive systems to advance Open Science (Buck 2021). A parallel of "performative dual-use risk mitigation" can be easily imagined, which would not necessarily be bad but likely still feature room for improvement. Stuart Buck's recommendation is to not only focus on promoting incentives for certain practices, but also to work towards a deeper culture change through

providing career incentives including strong signals from senior academics on hiring committees. These lessons seem transferable to dual-use risk mitigation.

The different challenges for implementing incentive systems demonstrate the importance of accounting for unfavourable effects which could reduce the value of panoptic dual-use management interventions. Costs of interventions can also generate overheads in terms of money, energy, and time, which could make research unjustifiably more costly. To be actionable and acceptable, costs of interventions need to be clearly lower than their benefits.

In the following, I evaluate different approaches for advancing panoptic dual-use management through incentive systems. First, I explore incentives for different stakeholders to reduce dual-use risks (section 12.2). Then, I take a closer look at the role of public perception and social pressure to reduce risks (12.3). I finish this chapter by looking at incentives for advancing low-risk alternatives and other risk-reducing research (12.4).

12.2 Incentive systems to reduce dual-use risks

Incentives to reduce dual-use risks may target different stakeholders and steps of the research life cycle. In the following, I present risk-reducing incentives for academic researchers, academic institutions, companies, and funding bodies. An overview of these stakeholders and their incentives is presented in Figure 12.1. I summarise and rate the promise of my proposals in Tables 12.1 and 12.2. I estimate the effectiveness of the proposed incentives for inducing risk-reducing behaviour. I evaluate how costly each proposal would be in terms of overheads, monetary cost, and loss in productivity. Lastly, I assess how feasible each proposal might be based on the acceptability and straightforwardness of its implementation.

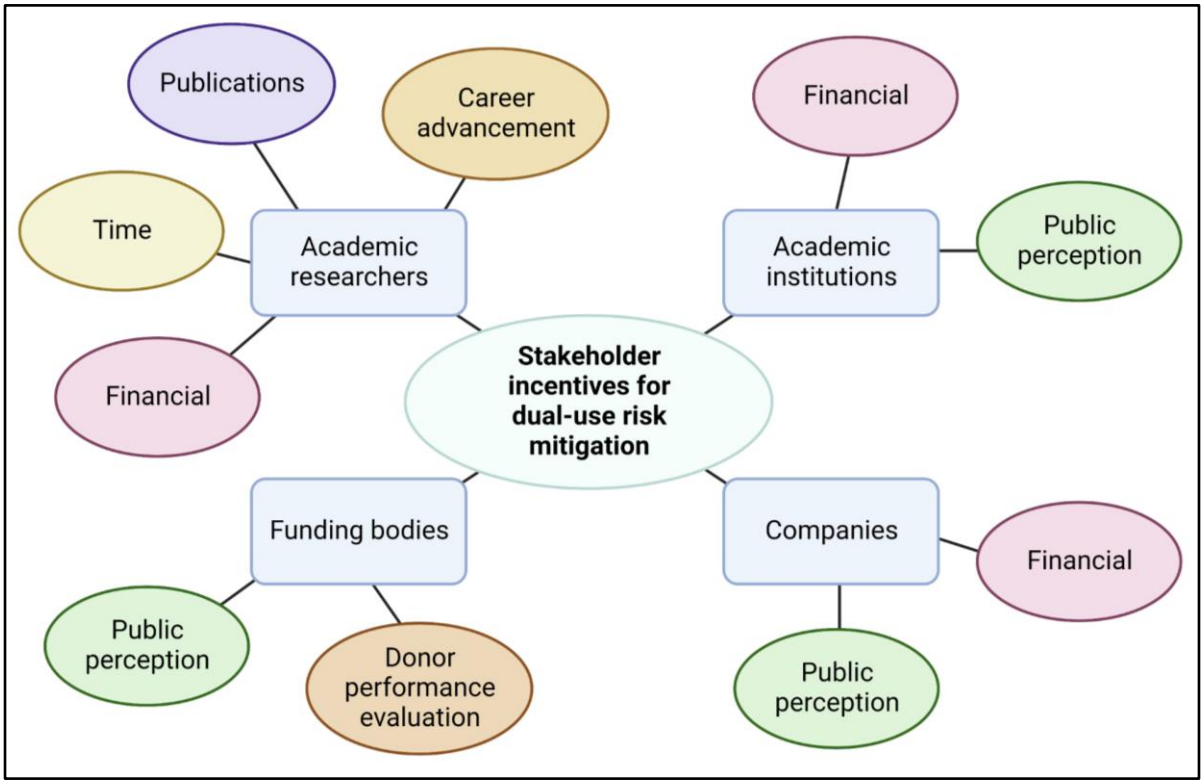


Figure 12.1: Stakeholder incentives for dual-use risk mitigation

Overview of crucial stakeholders and their incentives which could be shaped for dual-use risk mitigation.

12.2.1 Researchers

Targeting individual researchers with incentive systems might be very effective. Researchers sit at the core of the research ecosystem. They conceptualise new research, write grant applications, and conduct successfully funded projects. Academic research often forms the foundation for projects pursued by companies. Thus, if academic researchers preferentially pursue low-risk alternatives, this has trickle-down effects on commercial applications of biotechnology.

Academic researchers are usually not driven by financial gain but rather by factors for career advancement, like receiving grants and publishing high-profile articles. At the moment, researchers do not face many career incentives to minimise the dual-use potential of their research (and indeed, the opposite may apply). If dual-use risk was considered as part of grant funding or journal publication, this might encourage researchers to preferentially consider lower-risk research ideas. Dual-use risk mitigation may draw here on

lessons from efforts to promote Open Science. Barriers to advance Open Science values like equitable access and reproducible science are similar to barriers faced for dual-use risk mitigation. Researchers fear that commitment to Open Science and research transparency could hurt their career advancement (Schönbrodt 2016). Thus, the Open Science community has proposed different mechanisms for incentivising Open Science by considering it as part of career advancement factors - I draw on these as inspiration in the following sections.

One effective incentive might be to reduce the likelihood of receiving funding for dual-use projects. An inspiration for this could be efforts to promote "responsible research assessments", assessments of project proposals which include Open Science practices (Curry et al. 2020). Consideration of dual-use risks at funding bodies would incentivise researchers to think about these as part of project proposals. As the government is the source of most research funding, shaping funding decisions based on expected risks is an actionable and powerful lever. Ideally, funding bodies would discount research proposals proportional to their dual-use risk in funding decisions. I evaluate this idea more closely in Chapter 13. An alternative might be requirements for spending grant money. Such requirements could include spending a fraction of a grant on risk reduction, paying into a risk-reduction fund, or buying mandatory insurance. These requirements may be less effective incentives as increased costs may fall back onto funding bodies and thus may have a limited effect on what projects researchers choose to conduct.

Next to research funding, requirements that draw on researchers' time may also have incentive effects. Examples include requirements for paperwork, teaching, or training for researchers engaging in dual-use research. As discussed in the introduction to this chapter, researchers have noted that existing Select Agent and DURC regulations disincentivise such research (S. W. Evans et al. 2021). The substance of the paperwork or compulsory training may additionally reduce risks if it successfully motivates serious

engagement with risk-reducing practices. Increasing paperwork or training requirements for regulated research are generally widely practised and relatively accepted if accompanied by a good reason. However, depending on implementation, drawing on researchers' time may unreasonably reduce research productivity.

Lastly, publications might provide a potent lever due to their importance for academic career success. After the introduction of badges for journal papers that indicated compliance with Open Science practices, open data increased by more than order of magnitude (Kidwell et al. 2016). Similarly, badges could be considered for documenting critical evaluation of low-risk alternatives with a biosecurity or biosafety expert. However, as the majority of research is low risk, such badges might easily lose their currency. To tackle this, they could be reserved for documenting best standards in virological research.

However, such badges would not address the underlying problem that high-profile journals often incentivise the submission of controversial research, including if controversy arises from its risks (as discussed in Chapter 11, section 11.1.2) To create the most potent incentive for risk reduction, journals should stop providing incentives to conduct high risk research. Indeed, dual-use risks should negatively impact on the likelihood of a paper being accepted in a high profile journal. This seems feasible for at least the highest-risk research. In 2003, editors of top journals highlighted their consideration of dual-use risks in a joint statement, stating that under certain circumstances, they might not publish or redact dual-use research (Journal Editors 2003). It is important for high-impact journals to coordinate around publication decisions relating to dual-use risks, to prevent the Unilateralist's curse and ensure that no individual journal defects to leverage high-risk research for its citations that would boost a journal's impact factor. Journal editors could now come together to commit to not publishing research with a risk of catastrophic misuse.

Table 12.1: Incentives for researchers

Effect, cost, and feasibility of different measures are scored low, medium, or high based on qualitative assessment and consultation of experts.[48]

Strategy	**Effect**	**Cost**	**Feasibility**
Reduced likelihood of receiving funding for (some) dual-use research ● Requirement for external review (see Chapter 10) ● Dual-use risk considered during funding allocation decisions (see Chapter 13)	*High* Funding has a very strong lever on what research is conducted	*Low* May feature overhead and opportunity cost, however would be low for dual-use tiebreaker (see Chapter 13)	*Medium* Few downsides for government; Related systems already in place for human and animal subjects; Difficulty to create dual-use assessment that is robust and uncontroversial
Requirements for spending grant money ● On risk reduction activities ● Contribution to risk reduction fund ● Requirement to purchase insurance	*Medium* Not clear whether researchers will change projects, costs might simply be passed onto grantmaker	*Medium* Increases cost of any single project and associated with some overhead	*Medium* Attractive for researchers if risk reduction covered by funding, but then less effective; Costs (monetary, researcher time) reduce feasibility
Time-based incentives: Requirements for paperwork, teaching, or training for researchers engaging in (some) dual-use research	*Medium* Incentive effect potentially smaller than funding or publications; Risk-reducing training/paper-work could encourage risk mitigation and selection of lower risk projects	*Medium* Would take up researcher time; Government administration cost, especially for high-quality training; Teaching requirements would have lower costs	*Medium* Common practice, e.g. paperwork already in place for select agents; Government costs and overhead could be prohibitive

[48] Methodology used to assess effect, cost, and feasibility of interventions presented in tables 11.1-3: I used first principle-based analysis for initial evaluation. I explored additional ideas and corroborated uncertainties with experts including Elizabeth Cameron (academia; Brown University), Kevin Esvelt (academia; MIT), David Relman (academia; Stanford), James Diggans (industry; Twist Bioscience), Douglas Friedman (industry; BioMADE), Friederike Grosse-Holz (venture capital; Blue Horizon), Damien Soghoian (venture capital; GHIC). Discussions with these experts were unstructured and focused on topics relevant to their expertise. The assessments and the views they express (and any associated mistakes) are solely mine and not those of any of these experts.

High-profile journals agree to not to publish a subset of particularly concerning dual-use research (e.g. with significant potential for large-scale, catastrophic harm)	*Medium* Would in expectation only affect small but most important subset of research Many alternative publishing venues	*Low* Opportunity cost of not publishing high-risk articles; Coordination between journals needed	*Medium* Precedent exists for journal editors to make joint statement; less clear whether harder line for publishing is feasible

12.2.2 Academic institutions and liability insurance

Academic institutions are a crucial research stakeholder. They hire academic researchers, provide them with funding, and oversee regulated research through ethics and biosafety committees. Thus, academic institutions may be an important lever for reducing dual-use risks, even if they do not directly decide what research is conducted.

One proposed incentive mechanism targeted at academic institutions is mandatory liability and insurance for accidents or enabling misuse. Such a mechanism was discussed at a 2013 Wilton Park event (Wilton Park 2013) and proposed in more detail by Farquhar *et al.* (Farquhar, Cotton-Barratt, and Snyder-Beattie 2017). Mandatory liability and insurance could effectively internalise dual-use risks. Liability has a strong deterrent effect, driving applied safety and security research (Shavell 1984). As mentioned earlier, liability insurance is already used for small-probability high-consequence risks, for instance for accidents at commercial nuclear power plants, including melt-downs caused by terrorists (Berkovitz 1989; United States Senate Committee on Environment and Public Works 2002, 52). Insurance models have also been used to ensure against other deliberate events like terrorism (Woo 2002) or cyberattacks (Garrie and Mann 2014).

Different considerations could guide the implementation of mandatory liability insurance for dual-use risks. Governments could mandate liability for research risks as part of general liability insurance. Ideally,

liability would cover the whole spectrum of research risks. Still, if premiums in such a case are too large, a focus on catastrophe liability might be more actionable. Catastrophe liability could only trigger if an event exceeds a specified number of fatalities, such as one million (Esvelt 2022). Insurance companies could base premiums on whether an institution pursues research with pandemic-level risks and the existence of risk-mitigation measures. Liability insurance might work particularly well for biosafety risks. However, liability insurance may also work for foreseeable risks of misuse, such as misuse enabled by the enhancement of potential pandemic pathogens or the discovery of new pathogens. Indeed, market share liability in US law means that liability could even be feasible in cases where misuse cannot be traced back to any single piece of dual-use research. Market share liability was introduced in the seminal 1981 ruling *Sindell v. Abbott Laboratories*, where liability for reproductive cancers caused by prenatal exposure to the drug diethylstilbestrol (DES) was assigned to manufacturers according to their market share. Similarly, liability for misuse of research might be assigned across institutions producing enabling research.

Mandatory liability insurance for dual-use risks has multiple limitations, some of which may be overcome. First, pandemics are very costly, and insurers may not want to insure against catastrophes of such extent. Potentially enormous payouts were also an issue for insuring nuclear power plants until the 1957 Price-Anderson Act, which capped liability at an industry-wide figure (Farquhar, Cotton-Barratt, and Snyder-Beattie 2017; Berkovitz 1989). Second, quantifying the risks of deliberate misuse may be difficult due to uncertainty about malicious actors and their motivations and capabilities. However, as mentioned above, insurers cover terrorism in other areas (Woo 2002). Third, many institutions doing dangerous research are government-run and most dangerous research is government funded. As the government insures itself, there is less of a point in taking out insurance. The cost of a deliberate pandemic falls back onto the government itself. Despite this theoretical complication, governments may still care about getting incentives right for their research institutions and funding bodies. Additionally, liability insurance could

be first implemented for non-governmental research; this may be especially sensible given the dual-use risks of non-governmental research are currently governed the least well. Lastly, mandatory liability insurance may increase the cost of research. As mentioned above, premiums could be kept low by limiting payouts to the case of catastrophic events. Additionally, Farquhar *et al.* note that governments could increase life sciences budgets to compensate for increased research costs (Farquhar, Cotton-Barratt, and Snyder-Beattie 2017). Governments would thus shift unpredictable risk mitigation and catastrophe response costs to life sciences funding decision processes that reduce risks.

Another interesting incentive for academic institutions is public perception. Academic institutions rely on their public image to attract talented students, exceptional researchers, and donations. Thus, mandatory or voluntary disclosure of aggregated data on dual-use research at a given institution might provide a potent incentive to preferentially explore low-risk projects. Similar disclosure processes have been effective at reducing the carbon emissions of companies. I discuss the power of information as an incentive mechanism in more detail in section 12.3.

Table 12.2: Incentives for academic institutions, companies, and funding bodies

Effect, cost, and feasibility of different measures are scored low, medium, or high based on qualitative assessment and consultation of experts.[49]

Strategy	**Target**	**Effect**	**Cost**	**Feasibility**
Liability and requirement to buy insurance	Academic institutions, companies	*High* Financial internalisation of dual-use risks	*Medium* Makes research more expensive	*Medium* Precedent in related areas; Implementation challenge of cost and how to do risk evaluation
Public perception:	Academic	*Medium*	*Low*	*High*

[49] For methodology, see footnote introducing table 12.1.

disclosure of aggregate data on dual-use research	institutions, companies, funding bodies	Societal pressure may have limited incentive effect; Voluntary methods are limited	If infohazards can be managed limited cost; Low costs and only some staff time required	Precedent for decarbonisation; No significant barriers
Dual-use tax	Companies	*High* Financial internalisation of dual-use risk could be very effective; Effect on companies not clear	*Medium* Makes research more expensive	*Medium* Limited interest in barriers for biotechnology especially given competition with other countries
Risk-sensitive investors and divestment	Companies	*High* Very effective for start-ups, especially as market cool off	*Medium* Coordination across investors required; Only so much can be shaped form outside	*High* Easy for individual investors, difficulty lies in coordination and large scale adoption

12.2.3 Companies and taxes

Biotechnology companies are another target stakeholder for incentives to reduce dual-use risks. As gene and cancer therapy research increasingly advances viral engineering (see Chapters 5 and 6), private companies are becoming increasingly contributors to dual-use research. Unique factors need to be considered regarding mitigating dual-use risks at companies. Companies often spin out of academic research with a relatively narrow focus; therefore, companies may already be locked into certain methods and have less flexibility to switch to low risk-alternatives. At the same time, companies often distribute technologies widely. Thus, it might be particularly promising to incentivise companies to engage in biosecurity-by-design and structured access. Companies have frequently been the target of efforts to internalise externalities for carbon emissions. Thus, I particularly draw inspiration from such decarbonisation interventions and evaluate whether similar approaches could be feasible to incentivise the reduction of dual-use risks.

Carbon pricing, for instance through taxes, is the archetypical way of internalising negative externalities of carbon production (Kaufman 2016). Similarly, taxes could be used to price dual-use risks into company decisions. A tax might have similar incentive effects to liability and insurance - and either could apply to both companies and academic institutions. Even if a tax was passed onto customers, this could make dual-use products potentially less competitive. The upside of a tax versus a liability approach is that there is neither a need for involving insurers nor for a clear causal chain to misuse to assign liability. A tax would not necessarily have to be uniform but could be based on different tiers of risks. I provide an example of a tiered scoring framework for dual-use risks in Chapter 13. Similar to a tax, Farquhar *et al.* propose risk-proportionate payments based on centrally commissioned risk assessments (Farquhar, Cotton-Barratt, and Snyder-Beattie 2017). Governments could put the revenue of a dual-use tax towards funding concrete risk-reducing projects;[50] thus, governments could link a growing bioeconomy to investments for preventing catastrophic misuse of biotechnology.

While potentially an economically elegant idea, a tax-based system has limitations. First, carbon taxes have had a limited effect, partially because of political hurdles (Green 2021). Similarly, a pushback on a dual-use tax from companies and policymakers seems also likely, especially given increasing national competitiveness around the bioeconomy.[51] Second, if a national government imposed a dual-use tax, a potentially large fraction of relevant research would shift to other locations with less strict rules.[52] This

[50] Criteria for what constitutes a project that concretely reduces catastrophic misuse should be defined in advance; scope should be as narrow as possible to ensure risk-reducing effect, e.g. focused on interventions that directly prevent and reduce the impact of the misuse of specific technologies. An example could be the funding of national DNA synthesis screening.

[51] At the same time, if a government is willing to invest in the bioeconomy, targeted tax credits for risk-reducing projects could be a promising option. I discuss this idea in section 12.4.

[52] The downside of this may be exacerbated that a jurisdiction with lax rules in one domain (e.g. dual-use tax) might also have lax rules in other related biosafety and biosecurity domains. Thus, the risk of the research might increase if it is displaced to the jurisdiction with more lax rules.

could make a dual-use tax very unpopular among policymakers, especially if strengthening the national bioeconomy is a policy priority. However, evidence from carbon taxation shows that resulting increases in energy-intensive imports were very small and limited to select products (Dechezleprêtre and Sato 2017). Additionally, tariffs and other trade restrictions could be used to partially offset increased costs compared to foreign companies complying with more lax standards. For instance, the EU has long used tariffs for agricultural imports to protect local producers from foreign competitors that can produce at lower cost (Pigłowski 2021). However, tariffs seem less effective for knowledge-based research products. Third, risk assessments to inform taxation may be contentious and costly. Fourth, it may be difficult to assign taxes fairly based on the volume of dual-use research at a company. Taxes could be proportional to the number of individuals working on a project with dual-use risk. However, such a system may fail to capture companies relying on automation, which may produce similar or greater amounts of dual-use information. Lastly, taxes might be better at encouraging shifting to low-risk alternatives than inducing risk-reducing interventions. However, as mentioned in the introductory paragraph of this section, companies often focus on distributing technologies previously generated through academic research; thus, promoting risk-reducing interventions like biosecurity-by-design or structured access may be more important relative to incentivising low-risk alternatives for companies.

Requirements by funders and investors could be a potent incentive for companies to engage in dual-use risk reduction. Investors have become more interested in environmental, social, and governance (ESG) factors (Park and Jang 2021). ESG currently includes considerations of sustainability, social contribution, and ethical behaviour. ESG could similarly consider research risks, including dual-use potential. Even if no critical mass of investors agrees to consider dual-use risks as part of ESG assessments, individual investors may still make a difference. Investors already ask companies for a cybersecurity plan when considering investments (Douglas Friedman 2018). Similarly, investors could ask companies for a biosecurity plan,

raising awareness and inspiring consideration of dual-use risks. As part of investment contracts, investors may require risk-reducing activities, such as structured access and biosecurity-by-design, or discourage future exploration of high-risk projects. Douglas Friedman, when he was president of the Engineering Biology Research Consortium, proposed dedicating 1% of early biotechnology investments to biosecurity, including the mitigation of dual-use risks (Douglas Friedman 2018). Dedicated resources for biosecurity would at the very least help build a culture of biosecurity awareness..

Lastly, public perception and disclosure of information may also be an important incentive for companies. I look at this in section 12.3.

12.2.4 Funding bodies

Funding bodies might be a very effective target for incentive systems. Funding bodies sit upstream in the research lifecycle and decide which research to fund. In theory, governmental funding bodies are directly aligned with the government's goal of maximising benefits. In practice, staff at funding bodies may have their own incentives and it is members of the scientific community who are deciding which research proposals to fund. Thus, incentive systems for governmental grantmakers could still be valuable and effective.

Crucially, government performance evaluation of governmental funding bodies needs to consider efforts to mitigate dual-use risks. As funding bodies are tasked by governments to deploy taxpayer funds to advance research with the greatest overall benefit for the population, processes are necessary to track how effectively dual-use risks are managed.

Relatedly, disclosure of information and public input may be a crucial incentive for governmental funding bodies. In democratic systems, the budgets of governmental funding bodies are indirectly controlled by the voter. Thus, public perception may constitute a strong incentive for funding bodies. In previous sections, I mentioned that this non-financial incentive might also be effective for academic institutions and companies. Due to its cross-cutting relevance, I evaluate the disclosure of aggregate data on dual-use research in the following section.

12.3 Public perception and the disclosure of aggregate data on dual-use risks

Public scrutiny may be an important mechanism to reduce dual-use research. As discussed in Chapter 2, MacIntyre et al. found that 77% of interviewed members of the general public were unaware that dual-use research of concern is regularly conducted, and 64% deemed such research unacceptable or were unsure about its acceptability - with increasing information decreasing acceptance (MacIntyre et al. 2020). To allow society to respond to dual-use risks, disclosure of aggregated data on relevant activities is crucial. In systems theory, the power of information loops is well-recognised (Meadows 1999). Donella Meadows provides the example of the US Toxic Release Inventory requiring in 1986 the public disclosure of hazardous air pollutants released from factories (Meadows 1999). Within four years after mandated disclosure, Meadows claims, emissions had dropped by 40%.[53] Thus, disclosure of aggregated data on dual-use research might similarly incentivise reducing relevant publicly scrutinised activities.

One risk of publishing statistics on dual-use research is a potential public overreaction that could stifle relevant research to an unjustified degree. The above mentioned survey results by MacIntyre *et al.* demonstrate limited awareness and possible repudiation of dual-use research. However, in principle, the

[53] Meadows does not give a source for this number. Hanson gives a number of 26% for reduction in toxic chemical releases (Hanson 1992).

public should be informed on dual-use research given they are ultimately the ones affected by risks and benefits, and in the majority of cases help to fund research through their taxes. The solution to preventing an overreaction should not be to keep the public out of the loop, but rather education on dual-use risks and the importance of relevant research.

Another important downside of encouraging disclosure of dual-use research is the potential for drawing attention to hazardous information. Transparent disclosure of dual-use projects may point potential perpetrators to research with potential for misuse. A system for dual-use research disclosure needs to be designed with information and attention hazards in mind. To prevent such attention hazards, stakeholders could stick to public disclosure of aggregated data on dual-use research. For instance, institutions could disclose how many projects of different categories of risk were conducted. An example of possible risk categories are the tiered dual-use scores that I introduce in Chapter 13. Governments could follow-up with individual organisations to receive any additional details confidentially.

Stakeholders like funding bodies, companies, or academic institutions could disclose aggregate data on dual-use research voluntarily or because of mandates. An example of mandatory disclosure is the US Toxic Release Inventory discussed above. Stakeholders may self-disclose voluntarily because of incentives and public pressure. The Carbon Disclosure Project (CDP) operates in this way, a non-governmental and not-for-profit organisation that hosts a well-designed questionnaire to report environmental sustainability-related data. The CDP has successfully incentivised more sustainable business practices. Additionally, the collected data has enabled a better understanding of what interventions are effective at changing company behaviour (Green 2021; Ott and Endrikat 2022). Thus, data on dual-use research could enable the evaluation of dual-use management policies.

Disclosure of dual-use research may also be voluntary. A relevant example are the confidence building measures (CBMs) of the Biological Weapons Convention. As part of these CBMs, member states can publicly or privately disclose conducted research, such as publications of military biodefense laboratories (German Federal Foreign Office 2022). Disclosure of potentially dual-use biodefense research in this way has the main goal of building confidence in a country's compliance with the Biological Weapons Convention, but might also incentivise countries not to conduct research that may be construed as offensive. Because submission of CBMs is not enforceable and requires staff time, many countries are not regularly submitting CBMs - showcasing that where disclosure is voluntary, appropriate incentives are needed.

A model for effectively encouraging self-disclosure of dual-use research could combine mandates for government-sponsored research institutions and incentives for companies. Luo *et al.* found that disclosure to CDP was dependent on existing economic pressures, such as an emission trading scheme, as well as on social pressure, as bigger companies had a higher tendency to disclose (L. Luo, Lan, and Tang 2012). Similarly, economic and social pressure could be leveraged for voluntary disclosure of dual-use research. To create social pressure, news outlets, non-governmental organisations, and advocacy groups could appeal to stakeholders to fulfil their role in the social contract and engage in voluntary disclosure (Solomon and Lewis 2002). Lastly, self-disclosure would be greatly facilitated by universal, clear metrics of what to disclose. Such metrics could draw on existing methods, such as Select Agent, DURC, and P3CO definitions. Ideally, a tiered approach would be used, an example I outline in Chapter 13.

Independent of self-disclosure, information may also be collected by an external organisation based on publicly available material. An existing example of this model in biosecurity is the Global Health Security Index, which tracks countries' capabilities to prevent and respond to biological events (Nuclear Threat

Initiative and Johns Hopkins Center for Health Security 2021). An external organisation could scrape grants and publications for indicators of dual-use risk and publicise findings on the internet. In contrast, external information collection may be less effective at inducing companies to reduce dual-use research. However, an external model of establishing information loops might be easier to implement and still provide value, especially through providing data to inform other risk-reducing efforts.

12.4 Incentives and related strategies to advance risk-reducing projects

An important aspect of panoptic dual-use management is fostering the preferential advancement of risk-reducing projects. So far, I have mainly discussed interventions to disincentivise high-risk research. However, other strategies specifically aimed at advancing low-risk alternatives and other risk-reducing projects may uniquely contribute to panoptic dual-use management. Geels *et al.* have theorised that deep decarbonisation requires a combination of weakening existing systems, strengthening exogenous pressures to replace risk-increasing technology, and increasing the momentum of alternative niche innovations (Geels et al. 2017). I have already discussed parallels to the first two aspects of this three-pronged approach. In the following, I look at the last one: how to increase the momentum of risk-reducing innovations by incentivising low-risk alternatives and other risk-reducing projects.

Making low dual-use solutions to scientific problems more attractive could result in a positive feedback loop. Advancing low-risk approaches reduces dependence on high-risk approaches, which opens the possibility for greater regulation of high-risk approaches, which in turn inspires advancement of low-risk approaches. Importantly, I expect incentives for advancing risk-reducing projects to be more popular among researchers and policymakers than disincentives for dual-use research. Creating incentives for risk-reducing technologies does not only lie in the hands of the government. Philanthropists and other private grantmakers may use their grantmaking power to advance risk-reducing projects (see Tables 13.3 and

13.4). I first consider strategies to advance low-risk alternatives and then discuss strategies to incentivise other risk-reducing interventions like biosecurity-by-design and structured access.

Monetary incentives may be particularly effective at advancing low-risk solutions (Table 13.3). As discussed in Chapter 11, a well-defined strategy for advancing risk-reducing innovations is strategic niche management (Schot and Geels 2008). An example are the subsidies that helped achieve the commercial viability of photovoltaics (Schot and Geels 2008; Haegel et al. 2017). Similarly, increased funding of low-risk alternatives might increase research on relevant projects.

Early career grants may be particularly promising for advancing low-risk solutions as they will create an influx of talent. For funding dedicated to advancing new cancer or gene therapies, grants specifically for low-risk alternatives might be twice beneficial: they generate new insights and reduce the risk of misuse. Targeting early career researchers could be very effective at shaping the future trajectory of research in a field, say by setting a junior researcher up for a career researching non-viral gene delivery methods rather than viral vector research. This strategy may be particularly attractive to large governmental funders with a broad remit to advance science. However, for funding bodies that see reducing biological risks as their main objective, most such grants for advancing low-risk alternatives will likely not be cost-effective. This is because funding for low-risk alternatives, such as non-viral vector approaches for delivering therapeutics, only indirectly reduces misuse risks. The indirect effects of displacing high-risk technologies will generally be smaller than the effects of projects that directly reduce biological risks. A potentially more cost-effective strategy could be for biosecurity funders to give feed-in grants for low-risk alternatives, for instance grant programs to top up NIH grants meeting specific criteria. However, as humans generally value gains less than losses (Lee 2015), stocking up existing grants may thus be less effective at encouraging low-risk alternatives compared to slightly reducing the chance of a grant based on dual-use risks.

One other idea for using grant-making to preferentially advance lower risk alternatives is to allow rejected grants to be resubmitted with changes that reduce dual-use risks. This would require setting up a mechanism to judge whether a proposed features reduced risk compared to a previous submission, which may synergise well with my proposed for a tiered dual-use risk assessment in the next chapter. If implemented suboptimally, reductions in dual-use risks may mainly be superficial. However, if implemented in a way that requires substantial reductions in risks and use of new methodologies, this could incentivise researchers to consider new lower-risk research avenues.

Another strategy to foster innovation of low-risk technologies could be to make key intellectual property (IP) rights freely available. Governments, philanthropists, or other private funding bodies could buy relevant IP and licence it free of charge. As suggested in Chapter 11, lipid nanoparticle formulation is a limiting factor for the advancement of non-viral delivery of genetic materials to cells (Sandbrink and Shattock 2020). If a major philanthropist made the IP for lipid nanoparticle formulation publicly available, more academics and companies might start researching low-risk non-viral delivery methods. Especially given the success of RNA vaccines in the COVID-19 pandemic, there may be great interest in shifting towards RNA-based vaccines in academia - free IP would lower the barrier to entry.

Also non-monetary incentives could be used by different stakeholders to promote advancement of low-risk alternatives. For instance, universities and other research communities could promote low-risk solutions preferentially to young researchers. The iGEM competition has previously advocated for participating high school teams to preferentially use cell free technologies. iGEM has used blog posts, reports, and ambassadors to highlight how cell-free approaches are easy, safe, and innovative (Costa and

Hyde 2022).[54] In the early days of synthetic biology, the community advertised chances for fast career advancement given the novelty of the field (Delebecque and Philip 2015) - such career incentives could similarly be advertised for budding low-risk solutions.

Table 12.3: Strategies for advancing low-risk alternatives

Effect, cost, and feasibility of different measures are scored low, medium, or high based on qualitative assessment and consultation of experts.[55]

Strategy	**Actors**	**Effect**	**Cost**	**Feasibility**
Grant programs for risk-reducing technologies, in particular for early career researchers	Government, private grantmakers	*High* Shape the researcher pipeline to work on low-risk technologies	*Medium* Funding required; Potential opportunity costs of selective allocation	*High* Grantmakers could easily deploy funding in this way; Similar programs exist
Feed-in grants for low-risk alternatives, e.g. philanthropic funders top up relevant NIH grants	Government, private grantmakers	*Medium* Feed-in grants might not counterfactually change research	*Low* Relatively low additional research funding required	*High* Grantmakers could deploy funding in this way
Allow rejected grants to be resubmitted with changes that reduce dual-use risks	Government, private grantmakers	*High* Researchers would attempt to decrease dual-use risks, but may be mainly superficial	*Low* Little additional funding required, costs are associated with need to review more project proposals	*Medium* Would need to set up mechanism to rate risks; Complements dual-use tie breakers (Chapter 13)
Providing crucial intellectual property (IP) for low-risk alternatives for free, e.g. lipid nanoparticle formulation	Government, private grantmakers	*High* Could in select instances be highly effective at enabling academics to work on low-risk technologies	*Medium* Costs to buy IP may be significant	*Medium* Not clear whether cost-effective; Not clear whether companies would be willing to sell IP

[54] Personal communication with Dr Piers Millett, former VP for Responsibility at iGEM

[55] For methodology, see footnote introducing table 11.1.

Advertising promising low-risk technologies to young researchers, and highlighting career opportunities	Government, academic institutions, other research communities	*Medium* Effect more limited than that of research funding, but potentially still significant	*Low* Costs are mainly human resources for engagement work	*High* Accessible to many actors; Some versions already done

Table 12.4: Strategies for advancing risk-reducing practices, including biosecurity-by-design and structured access

Effect, cost, and feasibility of different measures are scored low, medium, or high based on qualitative assessment and consultation of experts.[56]

Strategy	**Actors**	**Effect**	**Cost**	**Feasibility**
Consideration of dual-use engagement as part of academic job interviews	Academic institutions, government	*High* Could create lasting change in scientific culture	*Low* Low costs	*High* Precedent in Open Science; Easy to implement, mainly awareness challenge
Government funders spend 5% of funding of synthetic virology on biosafety and biosecurity projects	Government	*High* Could build awareness and research ecosystem for reducing dual-use risks	*Medium* Significant research funding required	*Medium* Precedent for ELSI in human genome project
Tax benefits for companies to committing to certain rules, e.g. only using screened DNA	Government	*Medium* Could provide incentives for companies to think about biosecurity	*Medium* Potentially significant reduction in tax income	*Medium* Significant costs and challenging implementation details; Potentially highly feasible where interest in boosting bioeconomy
Mandate federally-funded research to use screened DNA products (Isaac	Government	*Medium* Could inspire more gene synthesis	*Low* Potentially need for small level of government	*High* Very feasible and accepted; Main challenge are

[56] For methodology, see footnote introducing table 12.1.

2022)		companies to screen	subsidy	implementation details
Mandating gene synthesis screening	Government	*High* Could very effectively incentivise screening, even beyond local companies	*Low* Potentially need for small level of government subsidy	*Medium* Some legal hurdles; Main challenge are implementation details

More generally, career incentives for academic researchers should be aligned to reward behaviour that mitigates dual-use risks. The Open Science community has considered how to shape career incentives to encourage researchers to engage in practices that increase transparency and reproducibility of research. One concrete mechanism that has been used is to ask researchers about their Open Science efforts as part of interviews for faculty positions (Schönbrodt 2016; Ulrich Dirnagl 2018; Buck 2021). Similarly, researchers could be asked for engagement with dual-use management and commitment to risk-reducing practices.

Governments and other grantmakers may fund research to directly advance risk-reducing interventions, including safety technologies like gene synthesis screening (see Table 13.4).. For instance, the philanthropic grantmaker Open Philanthropy is already funding a lot of dedicated biosecurity research, including explicitly giving early career grants - one such grant has funded this thesis (Open Philanthropy 2022). Government funders could fund risk-reducing disciplines proportional to risk-increasing work. For instance, funding bodies could allocate 5% of all funding for synthetic virology to biosafety and biosecurity projects. A precedent for this exists. Between 3% and 5% of human genome project funding was spent on characterising ethical, legal, and social implications (ELSI) (National Human Genome Research Institute 2012). The US government viewed studying ELSI as vital to the success of the human genome project. Similarly, biosafety and biosecurity are arguably vital to the success of synthetic virology. Studying how to

mitigate risks would contribute to proactive and proportional risk mitigation, while preempting heavy-handed regulations in the case of misuse.

Another potentially viable strategy for fostering risk-reducing behaviour are tax incentives, which have been used to incentivise carbon capture and sequestration (Anderson et al. 2021). For instance, tax credits could be given to biotechnology companies that follow a package of risk-reducing measures, like producing or buying screened DNA,[57] providing API-mediated access, or having a dedicated biosecurity plan and dedicated staff members. Early stage biotechnology companies frequently do not have any taxable revenue, in such cases, other forms of remuneration could be considered. Subsidies for companies conducting low-risk biotechnology research could be particularly attractive in the current political environment in the United States, where a high priority is given to strengthening the bioeconomy.

Governments could also use their procurement power to incentivise risk-reducing practices. For instance, research conducted at federal institutions could be required to use screened DNA (Isaac 2022). As discussed in Chapter 3, the US states of Maryland and California previously considered bills to mandate that state sponsored research only use screened DNA. While both original initiatives failed, a recent smaller California bill was successful and is now mandating that the California public universities only use screened gene synthesis products (California State Government 2022). One challenge with advancing regulations on gene synthesis screening is that defining standards and metrics is difficult and might inadvertently create lock-in effects resulting in worse outcomes than industry self-regulation.

[57] Discussed in Chapter 3, section 3.4.1. Screened DNA refers to gene synthesis products that have been checked to see if they encode for regulated pathogens or other dangerous proteins.

Next to strategic niche management, another technique for advancing niche innovations is technology forcing. A classic example of technology forcing are the California clean air standards of 1988 which required 2% of sold cars to be zero-emission vehicles (Schot and Rip 1997). Through these standards, the government forced car manufacturers to develop attractive electric vehicles. Similarly, the government could mandate gene synthesis screening or structured access to computational tools, which would induce the development of these risk-reducing technologies. Mandating US companies to only produce and use screened DNA would be very effective at establishing gene synthesis screening nationally - and screening standards might even spread internationally. The biggest hurdle to mandating gene synthesis screening is technical and financial: what does it mean to screen properly and who should pay for screening? Standardised tests and metrics are needed to identify a baseline that raises the *status quo*. Proving compliance with screening regulations might be costly, which could be prohibitive for smaller companies - thus the government might provide financial support, at the very least for testing compliance. Philanthropically funded free tools, such as SecureDNA, will also contribute to cutting screening costs ('Secure DNA Project' 2022).

The diverse set of stakeholders in the dual-use research enterprise offer many different angles for incentive-based interventions. In this chapter, I discussed a range of approaches, from shaping academic career incentives through research funding to using social pressure to induce organisations to reduce risky research. However, my discussion of the challenges of implementing incentive systems has highlighted that for aligning incentives with dual-use risk mitigation the devil is in the details: how to quantify and compare dual-use risks? In the next chapter, I sketch out how a concrete, simple-to-implement version of panoptic dual-use management may look and how to make the comparison of risks between project proposals actionable.

Chapter 13: A framework for comparative risk-benefit assessment

At the moment, researchers do not have the incentive to consider the minimisation of risks when conceptualising research. In Chapter 13, I argued that incentives could encourage researchers to preferentially consider low dual-use options when planning their research. One such incentive could be discounting grant proposals for dual-use research, so that research with low dual-use risks is funded slightly preferentially. This chapter discusses how to consider dual-use risks during funding decisions through comparative risk-benefit assessment.

In section 13.1, I define and motivate comparative risk-benefit assessment. In the subsequent sections, I sketch out its implementation. In section 13.2, I propose to use dual-use risks as a tiebreaker between projects of similar promise. In section 13.3, I develop a tiered scoring system to classify a spectrum of dual-use risks of virological research. I finish by calling for pilots of new governance methods (section 13.4).

13.1 Motivating comparative risk-benefit assessment

Currently, funding bodies like the NIH do not consider dual-use risks when assessing the potential promise and positive impact of a project. At the NIH, scientific review groups assign grant proposals an "overall impact score" based on their scientific promise (National Institute of Allergy and Infectious Diseases 2020). This "overall impact score" is informed by five separately scored criteria, which either describe a proposal's importance (Significance, Investigator) or likelihood of success (Innovation, Approach, Environment). While NIH regulations require this process to consider risks to human subjects, vertebrate animals, and select agents, the process does not formally consider dual-use risks. As I presented in Chapter 10, where existing policies require consideration of dual-use risks, dual-use risks are assessed for individual projects largely independent from funding decisions. Even for research captured by the DURC and P3CO policies, risks are only considered before or after a decision on whether a project should be

considered for funding. Risks are not factored into assessing the merit of a proposal compared to others. Thus, "overall impact scores" which decide which projects are funded, actually do not reflect a project's overall impact. The scores do not consider dual-use risks, despite these constituting possible negative effects of a project that decrease its expected impact.

Considering dual-use risks during funding decisions would make sense from an ethical perspective. First, considering both benefits and risks during funding decisions would maximise the expected value of funded research. Additionally, considering risks during funding decisions would impose minimally on academic freedom. Funding decisions generally aim at maximising the societal impact of grants and, therefore, already constrain academic freedom significantly (N. G. Evans 2018).

In the following, I look closer at how risks could be considered next to benefits when deciding what research to conduct or fund. I term the comparison of proposals across benefits and risks "comparative risk-benefit assessment".

> ***Comparative risk-benefit assessment:*** *Consideration of benefits and risks in decisions between research projects.*

Comparative risk-benefit assessment is one strategy that could help realise panoptic dual-use management. Consideration of risks would shift the research portfolio towards lower-risk alternatives - directly through funding decisions and incentivising researchers to submit low-risk proposals. Considering benefits and risks during funding decisions can identify alternative projects with roughly similar levels of benefits but substantially lower levels of risks. Thus, comparative risk-benefit assessment may enable the preferential advancement of low-risk alternatives for synthetic virology research. For instance, funding bodies might

preferentially advance low-risk non-heritable viral enhancement relative to similarly promising heritable approaches with greater dual-use risks.

Comparative risk-benefit assessment may be made actionable by focusing on relative risks and benefits comparisons. A challenge for risk-benefit assessment is the difficulty of quantifying absolute risks and benefits and comparing the two (Gryphon Scientific 2016). Few feedback loops exist for interventions aimed at preventing low-probability, high-consequence misuse threats, and risks are intrinsically dependent on human factors, including whether motivated malicious actors exist. However, relative comparisons of benefits and risks may be easier. Funding bodies already compare the relative benefits of research proposals. Similarly, relative risk assessments seem generally possible - I have used relative risk assessments throughout this thesis, for instance, when evaluating risks of different technologies in Part II. In the following, I present options for implementing comparative risk-benefit assessment and leveraging relative comparisons of risks and benefits to advance research with the greatest expected value.

13.2 Routes to implementation

Comparative risk-benefit assessment might reduce dual-use risks across the research portfolio if implemented at governmental funding bodies. Governmental research funders have the task of benefitting taxpayers and securing future economic growth. Part of this task is mitigating risks to society. Furthermore, funding bodies are upstream in the research lifecycle and shape what areas scientists are trained in or investigate. Funding bodies receive many different research proposals, many more than they can fund. Thus, opportunity costs to not fund a given proposal are low. Funding bodies already have established practices to engage in the difficult task of comparing the relative merits of proposals. Thus, governmental funding bodies are a promising site for comparative risk-benefit assessment. Nevertheless,

researchers can also apply comparative risk-benefit assessment when conceptualising research - and will be incentivised to do so, if funding bodies lead the way.

13.2.1 Risk scoring by scientific review groups

Funding bodies could assign comparative risk-benefit assessment to scientific review groups. Scientific review groups consist of scientific experts who already score the benefits of proposals. Ideally, these same groups would also rate risks and factor these into overall impact assessments and funding decisions. Reviewers need to consider protection for human subjects, ethical use of vertebrate animals, and "inclusion of women, minorities, and individuals across the lifespan in studied populations" (National Institute of Allergy and Infectious Diseases 2021). However, in practice, these factors are usually only considered superficially and in a tick-box format.[58] In contrast to the different aspects of scientific merit, factors like the protection for human subjects do not have explicit subcomponent scores. Thus, if scientific review groups are tasked with considering dual-use risks, dual-use risks should be assigned an explicit subcomponent score.

There are additional hurdles to scientific-review group-led comparative risk-benefit assessment. First, scientific review groups may lack the right expertise for risk assessment. While scientists on these committees hold part of the relevant expertise, they usually have had little training in ethics, security, and biorisk management. Thus, members with those backgrounds would need to be added to scientific review groups or scientific members would have to be trained in risk assessment. Second, there is a risk of information hazards being identified throughout the risk assessment process that are difficult to contain. Where scientists are trained in identifying relevant dual-use research, this could also result in spreading

[58] Personal communication with David Relman, based on his personal experience of serving on NIH Scientific Review Groups.

information hazards. Third, given the difficulties in absolute comparisons of risks and benefits, it may be challenging to find a consistent way to factor a risk subscore into overall impact scores. The following proposal for dual-use tiebreakers may circumvent this challenge.

13.2.2 Dual-use tiebreakers

One concrete way to realise comparative risk-benefit assessment would be to use dual-use risks as a tiebreaker in funding decisions. Using dual-use risks as a tiebreaker between projects on the border of being funded would mean that differences in the level of dual-use risks between projects could sway decisions. For example, if multiple proposals are considered a "maybe" for funding, those proposals meeting a pre-specified dual-use definition could be discounted (see Figure 13.1). Relevant dual-use levels need to be assigned by a specified process and could draw on a tiered scoring framework, such as the one I propose in section 13.3. I provide two concrete proposals for how a dual-use tiebreaker could be implemented at the NIH in Box 13.1.

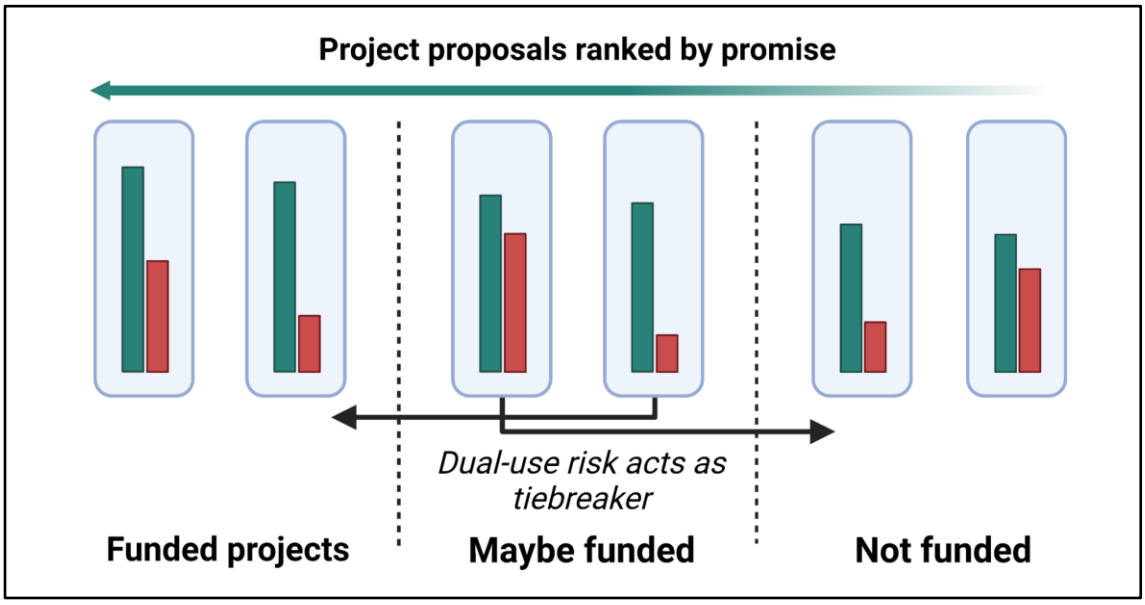


Figure 13.1: Schematic of a dual-use tiebreaker

Funding decisions between research projects with moderate promise (green), that funding bodies are uncertain about, are decided by associated dual-use risks (red).

Box 13.1: Concrete proposals for implementing a dual-use tiebreaker at NIH

Approach 1: Dual-use risk scores decide between borderline projects

Based on overall impact scores, project proposals are ranked. Depending on available funding, the top bracket of projects is funded. For instance, for the fiscal year 2023, NIH plans to fund the top 10% of established investigator grant applications (R01). Project proposals with scores around this 10% cut-off, say project proposals with scores ranging from 9.5-10.5%, could be compared for dual-use risks. Between these projects estimated to feature roughly similar promise, how significant associated dual-use risks are could decide on which project was funded. Projects could be assigned dual-use scores for different levels of risk, for instance with the framework that I introduce in section 13.3. Then projects could be ranked by dual-use scores and then the lower scoring half (or other fraction) funded.

Advantage: No need to directly trade off risks and benefits, dual-use risks are only compared between projects with very similar estimated levels of benefit.
Disadvantage: Requires a separate ranking, which could be perceived as substantial change to the funding decision process.

Approach 2: A dual-use modifier to discount project with higher dual-use potential

After reviewers have assigned an overall impact score from 1 (best score) to 9 (worst score), the overall impact scores of individual reviewers are averaged (National Institute of Allergy and Infectious Diseases 2020). At this stage, a modifier based on dual-use level could be added to each project's average score. For projects with the highest dual-use level, the highest modifier (e.g. +0.2) would be added to the averaged impact score. For lower dual-use levels, lower modifiers would be added. For the dual-use levels (DUL) 1-5 defined in section 13.3, example modifiers could be +0.2 for DUL5, +0.15 for DUL4, +0.1 for DUL3, +0.05 for DUL2, and no modifier for DUL1. Averaged scores with added modifiers would then be used as currently: scores would be rounded to one decimal place and multiplied by ten for the final score, which then decides which projects are funded. For example, the average reviewer score of a project is 2.34. Assume it involves creating a new synthesis protocol for horsepox and thus has dual-use level 4 (see section 13.3). Therefore, a modifier of +0.15 is added for a modified score of 2.49. The resulting modified score is rounded to 2.5 and multiplied by 10 for a final score of 25. Thus, dual-use risks modified the project's score from 23 to 25. For the fiscal year 2023, NIH plans to fund the top 10 percent of established investigator grant applications (R01). If adjusting for dual-use risk pushes a given

project outside of the top 10 percent of scored projects, then a different high-promise project will be funded instead. In theory, adding a dual-use modifier goes beyond a dual-use tiebreaker as it is applied to all projects. However, in practice, a slight modification of the impact score will only impact the funding outcome for projects at the brink of being funded; thus, this dual-use modifier could be considered a tie-breaking approach.

Advantages: Only existing ranking of overall impact scores required. Potentially greater incentive effect as dual-use modifier applies across the whole range of projects, even if funding outcome similar to Approach 1.
Disadvantage: Requirement to define a set a discount factor that is proportional to dual-use risks, which may be challenging and could be associated with controversy.

There are distinct advantages of implementing comparative risk-benefit assessment through a dual-use tiebreaker. First, as mentioned above, the proposed dual-use tiebreaker circumvents the difficulty of weighing off absolute risks and benefits. It simply requires a comparison of relative risks. Nevertheless, a dual-use tiebreaker may be an effective example of panoptic dual-use management. The dual-use tiebreaker would both directly cut dual-use risks and incentivise researchers to preferentially propose lower risk projects. Second, funding bodies may implement dual-use tiebreakers with minimal changes to the existing scientific review group process (see proposals for NIH in Box 13.1). This is especially the case if dual-use levels are assigned based on a predetermined framework - I expand on this in section 13.3.

A dual-use tiebreaker needs to be proportional. Policymakers need to balance dual-use related incentives with unjustified reductions in research productivity, such as through funding less promising projects. The size of dual-use tiebreakers can be finetuned by adjusting how many projects are considered as part of the tiebreaker (for Approach 1 introduced in Box 13.1) or the size of the dual-use modifier (Approach 2). Effects could be monitored to enable continuous learning around definitions of dual-use risks and the optimal size of the dual-use tiebreaker. Generally, it may be preferable to err on smaller modifications of

funding outcomes as long as they still effectively incentivise researchers to propose low dual-use projects preferentially. If the effective evaluation of the benefits of projects is difficult or there are many more high-promise projects than funding, the size of a dual-use tiebreaker may be increased. In such cases, there is less of a risk of shifting the portfolio to substantially less promising projects.

Implementing a dual-use tiebreaker is not a replacement for individual risk-benefit assessment of high-risk projects. As a dual-use tiebreaker relies on the relative comparison of risks and dual-use score modifiers are generally not proportional but smaller than risks, this mechanism won't consistently prevent the funding of high-benefit projects for which risks outweigh benefits. Thus, a dual-use tiebreaker complements individual project oversight. As discussed in Chapter 10, examples of individual project oversight include existing P3CO and DURC practices. However, ideally, an external, funding-body independent governmental committee would review high-risk proposals to prevent or modify projects with unjustified risks.

13.3 Tiered dual-use scoring of synthetic virology

A clear framework for dual-use scoring is important for dual-use tiebreakers and other panoptic dual-use management approaches. Panoptic dual-use management approaches could create adverse incentives to downplay the dual-use risks of their research. Ideally, experts would assign dual-use scores based on in-depth review and consideration of individual projects. However, in practice, a framework is needed that allows relatively objective scoring of different research proposals. Ideally, researchers should be able to apply it themselves to tick relevant boxes on their grant proposals based on what viruses and types of modifications their work involves. This self-categorisation could then be cross-checked during the compliance review. If the research appeared to be higher risk than assigned based on the scoring framework, NIH staff could forward the proposal to an additional, ideally external, dual-use review. A

framework that researchers can apply themselves will necessarily look more list-like, and will feature some of the limitations discussed in Chapter 8. For instance, it will offer limited resolution and might miss novel risks. Therefore, a tiered dual-use framework does not replace efforts to monitor and identify emerging risks.

A scoring framework should capture differences in dual-use risks. A tiered approach would allow going beyond the highest-risk research and also consider transfer risks arising from synthetic virology research for cancer and gene therapies. These increasingly popular research areas may not involve pathogen research but can still advance dual-use methods to enhance potential pandemic pathogens (see Chapter 9). Tiered approaches already exist in other areas of life sciences governance. For instance, biosafety requirements have historically been scored from Biosafety Levels 1 to 4, levels which define requirements for biocontainment and training (Ta, Gosa, and Nathanson 2018). Similar to how different levels of safety risks are considered when deciding what facilities are required, different levels of dual-use risks could be considered when deciding whether to conduct an experiment. In their most recent Laboratory Biosafety Manual, the WHO is now moving towards a risk-based approach to manage safety risks (World Health Organisation 2020). Similarly, dual-use risk assessment could also move beyond a set scoring system, once such a scoring framework has socialised dual-use risk mitigation.

In 2007, researchers at the University of Maryland proposed a tiered system for biosecurity governance (Steinbruner et al. 2007). They defined three categories of research of concern and proposed respective oversight structures. They proposed establishing international oversight for "activities of extreme concern, " which includes enhancement of listed high-risk pathogens.[59] They proposed a national oversight mechanism for "activities of moderate concern, " including enhancement of related agents. Lastly, the

[59] I discuss challenges of international coordination on dual-use risks in more detail in Chapter 14.

authors proposed local oversight for "activities of potential concern, " including enhancements of non-listed or related agents. Thus, with this tiered framework, the authors captured a significant range of transfer risks. These include transferable insights within viral families and other potentially transferable enhancements. I build on this University of Maryland framework in my own tiered risk-scoring framework.

13.3.1 A framework covering five dual-use levels

I propose defining five dual-use levels (DUL1-5) for research involving viruses (Table 13.1). This framework aims to capture dual-use risks that may enable the release of a pandemic-capable pathogen. It applies to classical virological research and other synthetic virology research, including viral vectors. Each of the dual-use levels features an underlying principle of research it covers: I define the highest-risk research, DUL5, as research that may result in the identification of a potential pandemic pathogen (PPP) or the enhancement of a PPP; both would add a new agent capable of causing massive harm to the arsenals of possible perpetrators. I define DUL4 as research that may directly enable the misuse of a known PPP or directly enable viral enhancement that may create a PPP. DUL3 captures research with similar but more indirectly transferable risks. DUL2 captures any work that advances general skills for creating or testing PPPs. DUL1 captures the lowest risk research, other research that does not involve self-replicating viruses. Based on these principles for defining DUL1-5, I define categories of research activities that fall into a respective dual-use level (Table 13.1).

For the purpose of this framework, I define a PPP as a pathogen that is likely to be highly transmissible and capable of causing significant harm in humans.[60] This includes pathogens with moderate transmissibility,

[60] PPP and ePPP definitions are adapted from the P3CO policy. U.S. Department of Health and Human Services. Framework for Guiding Funding Decisions about Proposed Research Involving Enhanced Potential Pandemic Pathogens. 2017 https://www.phe.gov/s3/dualuse/Documents/P3CO.pdf

such as smallpox, and pathogens with moderate virulence like SARS-CoV-2 (age dependent case fatality rate around or below single digit percentages). A PPP is also defined by a lack of pre-existing immunity in a population, thus enhancing an endemic virus to evade natural or vaccine-induced immunity may also result in a PPP. In contrast to PPPs, pandemic pathogens that are actively circulating in the population feature less dual-use risks, as an additional release would be substantially less harmful and there may be very great public health importance of characterising them. For the purpose of dual-use scoring, actively circulating (pandemic) pathogens may be treated as "agents related to PPPs".

Illustrative lists could further specify what exactly is captured under each research category used to specify the principles for each dual-use level. For instance, DUL4 includes the category of research "heritable enhancement of agents related to a PPP". An associated illustrative list might highlight enhancements of transmissibility, virulence, host range, or immune evasion on viruses of viral families with known potential pandemic pathogens, such as coronaviridae or orthomyxoviridae (includes influenza A virus). The DUL4 category "Work on enhancement methods applicable to creation of an ePPP" may be illustrated by a list that includes enhancements of transmissibility or immune evasion of paramyxoviridae, which may be applicable to Nipah and Measles virus, respectively, as well as universal genetic elements for immune evasion discussed in Chapter 6.

Table 13.1: Proposed dual-use scoring framework for virological research

Enhancement refers to the heritable enhancement of transmissibility, host range, virulence, immune evasion, or countermeasure resistance. Related to refers to viruses in the same viral family. PPP: potential pandemic pathogens. For the purpose of dual-use scoring, actively circulating (pandemic) pathogens may be treated as "agents related to PPPs".

Dual-use level (DUL)	**Virological research/synthetic virology research**

5	Research that may result in the identification of a PPP or enhancement that may create a PPP ● Enhancement of any agent that may create a PPP ● Work that may result in enhancement of a PPP ● Work that may result in identification of a new PPP
4	Research that may result in directly applicable methods for a) enhancement that could create a PPP or b) the creation/dispersal of a PPP ● Other work on a PPP ● Heritable enhancement of agents related to a PPP ● Work on enhancement methods applicable to creation of a PPP ● *In silico* design or identification of a PPP or an enhanced PPP ● Development of methods enabling synthesis of a PPP ● AI lab assistants for instructing synthesis of a PPP ● Aerosol production or dispersal of a PPP or related agents
3	Research that may result in indirectly applicable methods for a) enhancement that could create a PPP or b) the creation/dispersal of a PPP ● Any work on agents related to a PPP ● Heritable enhancements applicable to potentially autonomously transmissible viruses ● Aerosol production or dispersal of any virus ● AI lab assistants for instructing viral synthesis
2	Research that builds general skills for creating or testing PPPs ● Testing of any virus in animals or humans ● Work on replication competent-viruses *in vitro* ● Work involving *de novo* synthesis of any virus
1	Virological research with the least potential for misuse ● Other research including work on non-PPP-related viral proteins, or work on RNA or DNA-based gene delivery

13.3.2 Example applications

This tiered scoring framework can be applied to examples of research discussed in this thesis. Enhancing H5N1 avian influenza through its passaging in ferrets, as published by Fouchier in 2012 (Herfst et al. 2012), may result in the creation of a PPP; thus, it would be an example of DUL5. The Boston University experiments involving the insertion of the Omicron spike into the wild type SARS-CoV-2 backbone, discussed in Chapters 10 and 11, also feature the possibility of creating an enhanced pathogen (Chen et al. 2023). However, given SARS-CoV-2 is already circulating in humans, misuse potential is reduced

(partially because the virus is already circulating, partially because of natural and vaccine induced immunity). Therefore, these experiments would be treated as "enhancement of an agent related to a PPP" and thus DUL4. In the study by Chan *et al.,* repeatedly discussed throughout this thesis, the authors try to develop immune modulating genetic elements for improving AAV as a gene therapy vector (Chan et al. 2021). The design of universal immune-evading genetic elements could score between DUL3 and DUL4, because such elements could be transferable to PPPs. Because the elements by Chan *et al.* would likely not be sufficiently effective to significantly enhance a PPP, I would rank this study as DUL3. In contrast, AAV-specific immune evasion enhancements would score as DUL2, if it would involve the synthesis of virus or its testing in animal models. Work only involving the study of AAV capsid proteins in a bacterial expression system would qualify as DUL1; as would the vast majority of research on mRNA or DNA as gene delivery systems or vaccines.

13.3.3 Challenges and application

One challenge for creating a tiered dual-use scoring system is the objective ranking of dual-use risks. As discussed in Chapter 2, even biosecurity experts may disagree substantially on dual-use risks (Boddie et al. 2015). However, within a given topic area, such as virology, ranking categories of research by dual-use risks may be surmountable. I expect most experts to agree that creating a PPP through pathogen enhancement features greater dual-use potential than using a similar enhancement method on a related agent, which in turn features greater dual-use potential than virus-specific enhancement of a non-related agent. However, experts may still disagree about what experiments specifically fit each of these categories. New approaches for consensus-finding may help create dual-use frameworks with broad support. For example, the consensus-finding platform Polis allows users to vote on suggestions and uses a machine learning algorithm to identify points of agreement ('Polis' 2022).

One important question is whether adoption of risk-reducing measures should reduce the dual-use score. This makes sense in cases where dual-use risks are tangibly reduced. An example would be if a research project to discover new potential pandemic pathogens would commit to not publishing the genomes of discovered pathogens and only selectively sharing them with countermeasure developers. In such a case, the dual-use level could justifiably be reduced by 1 (or in rare cases even 2). In practice, relevant tangible risk reductions may not be possible for the fast majority of projects without moving away from the established model of sharing research results freely. For instance, a commitment to "tailored communication" would not warrant lowering of the DUL score, if tailored communication was limited to avoiding the highlighting of misuse potential during publication. For misuse risks that are more indirect, i.e. DUL2-3, reducing dual-use risks may be particularly difficult, as it is less one specific bit of information (e.g. a genome) that generates the risk but rather a combination of skills, ideas, and more diffuse knowledge on what engineering approaches are promising.

Science policymakers could adopt similar tiered dual-use frameworks for research areas beyond synthetic virology. Ideally, such frameworks would similarly use the DUL1-5 scale to allow relative comparison of risk levels across subject areas. Thus, research areas with less dual-use risk might only span DUL1 to DUL3.

The proposed tiered risk scoring framework has many applications beyond dual-use tiebreakers. Funding bodies may use this scoring framework for other implementations of comparative risk-benefit assessment. For instance, if scientific review groups assigned risk subscores, dual-use scores could feed into this process. The dual-use levels defined in this framework may also be used to decide whether a project is subject to individual project oversight. For instance, high-risk research that goes to an external risk-benefit assessment could be defined as DUL4 and DUL5, while a broader range of dual-use research from DUL3 could be evaluated at an institutional level; the resulting review process can help researchers identify risk-reducing

measures and possible changes to dual-use level throughout the project. Lastly, the tiered scoring framework could also be used for other panoptic dual-use management approaches. For instance, institutions could disclose how many projects of each dual-use level they support or dual-use risk scores could be used to guide premiums as part of a mandatory liability insurance scheme.

13.4 Piloting new governance approaches

In this chapter, I have argued for striving towards comparative risk-benefit assessment to increase the expected value of funded research and create incentives to reduce dual-use risks. I sketched out one concrete, actionable proposal in the form of dual-use tiebreakers and presented a tiered scoring framework that funding bodies could apply. Based on my assessment, implementing dual-use tiebreakers at governmental funding bodies could be an effective panoptic dual-use management method with limited costs and other downsides. However, theory only goes so far. Real-world experimentation with dual-use tiebreakers and other novel governance methods is needed (S. W. Evans et al. 2020). I believe dual-use tiebreakers could be an acceptable starting point for piloting comparative risk-benefit assessment. When governmental funders are unwilling to experiment with biosecurity governance, philanthropic funding bodies could step up and lead the way.

Piloting a range of approaches is a crucial aspect of advancing dual-use management. In the final part of this thesis, I draw all of my proposals together and discuss that it is not a single governance regime but rather a combination of governance approaches that is needed to mitigate risks. I highlight overarching challenges for advancing dual-use risk mitigation, such as uncertainty around risks, apathy to dual-use risks in the broader scientific community, societal input into dual-use management, and international coordination. Considering emerging threats, opportunities for dual-use management, and advances in

technologies to prevent and defend against pandemics, I try to chart an actionable way to a future with less pandemics.

PART V: CONCLUSIONS

Chapter 14: Beyond panoptic dual-use management

14.1 Introduction

In this thesis, I have analysed emerging dual-use risks in synthetic virology and their mitigation through different governance approaches. In the introduction to this thesis, I discussed the example of Chan *et al.*, a study that involved the generation of genetic elements for immune modulation. These genetic elements could be directly transferable and be used to enhance a large range of pathogens; nevertheless, no mechanism to reduce dual-use risks of this study existed. The Chan *et al.* study exemplifies my main contention of this thesis: existing governance mechanisms fail to address important sources of dual-use risks. In response, I consider actionable approaches for managing both the highest-risk research and a broader spectrum of dual-use research. I propose going beyond the governance of individual projects and to take a "panoptic" approach to reduce dual-use risks by creating incentives for stakeholders to consider dual-use risks when deciding between projects. In this chapter, I draw these different pieces together, highlight how they can be made actionable, and discuss overarching challenges to advancing dual-use management. I end with an outlook on how the proposed dual-use management approaches feed into the higher-level strategy for mitigating biotechnology misuse and, eventually, ending biological risks.

14.2 Dual-use management

Dual-use management describes strategies to mitigate misuse risks of dual-use research. History demonstrates that biotechnology misuse is a real risk. The Japanese doomsday cult Aum Shinrikyo tried to release anthrax in the 1990s; with modern technology, Aum Shinrikyo might have been able to create and engineer viruses for deliberate release. Dual-use management aims to constrain the capabilities of actors with limited resources to make the misuse of biology unattractive or unsuccessful.

Serious discussions about how to govern dual-use risks started in the United States in the 2000s. Anthrax attacks through the US postal system showcased the risks of bioterrorism, and high-profile publications highlighted unfettered advances in dual-use research. Over the last two decades, the United States has established several dual-use policies. At least since the high-profile controversy about "gain-of-function" experiments on avian influenza in 2011, other countries also have started to consider how to address dual-use risks from biotechnology.

Existing policies have been reactive and shaped by an influential scientific effort to retain self-governance. Because policies have been reactive to specific instances of dual-use research and have been drafted mainly by scientists, there has been relatively little comprehensive and interdisciplinary consideration of how to govern dual-use risks - despite lessons from other governance efforts potentially being applicable. An interdisciplinary approach is needed to make progress on forward-looking dual-use management. Such an approach requires understanding research and its risks and how to effectively impose policies based on sociological, economic, and ethical perspectives.

Dual-use management must not stand still - it needs to adapt to scientific advances that may create new misuse risks. One example is the area of synthetic virology, in which increasingly sophisticated and accessible methods to engineer viruses are advanced for many different purposes. Advances in synthetic virology may enable an increasing number of individuals or small groups to engineer viruses to do harm, which could result in potentially catastrophic pandemics significantly worse than COVID-19. Thus, a core part of pandemic prevention is preemptively addressing the dual-use risks of synthetic virology.

14.3 Themes and proposals

The starting point for my analysis formed the question of where and what dual-use risks are emerging from synthetic virology research. In Part II of this thesis, I evaluated the dual-use risks of different research areas currently not subject to dual-use oversight in the United States. I identify multiple key take-home messages.

In Chapter 4, I found that efforts to identify potential pandemic pathogens in wild animals feature high dual-use risks. Malicious actors may exploit identified pathogens to start or threaten to start a pandemic. Misuse risks of identifying new potential pandemic pathogens resemble those of the enhancement of potential pandemic pathogens, research generally seen as featuring the greatest risks. In the wake of the COVID-19 pandemic, more wildlife virus discovery efforts to identify potential zoonotic viruses may take place, efforts for which benefits may not necessarily outweigh risks.

In Chapters 5 and 6, I analysed emerging dual-use risks from using synthetic viruses for vaccines and gene therapy. I identify lines of research with significant dual-use risks, such as research on immune evasion approaches that are transferable to potential pandemic pathogens. These dual-use risks might increase as researchers increasingly develop platform approaches and universally applicable methods for viral enhancements. Examples are universal genetic elements for immune modulation, such as the Chan *et al.* study, and computational approaches for optimising viral vectors (Chan et al. 2021; Ogden et al. 2019). I also identified a theme with important implications for governance: there are frequently low-risk solutions with benefits very similar to those of higher-risk approaches. For instance, mRNA vaccines are a promising low-risk alternative to viral vector vaccines, and virus-specific modifications for gene therapy delivery may be as effective as universally-applicable viral enhancements.

In Chapter 7, I analyse impacts of artificial intelligence on biosecurity risks. I differentiate between the risks of large language models (LLMs) and biodesign tools. While LLMs may in particular increase the accessibility of biological agents, the latter might in particular increase the ceiling of capabilities. Computational tools empowered by artificial intelligence highlight the importance of general-purpose technologies for biological misuse, which also includes DNA synthesis. Throughout Chapters 4-6 I also identified how more specialised computational tools empower a wide range of dual-use capabilities, including predicting the chance that a wildlife virus might infect humans and optimising viral vectors across multiple properties.

The second main section of my analysis put emerging dual-use risks into the context of existing dual-use management approaches. In Part III, I identified how dual-use oversight in the United States fails to keep ahead of risks from synthetic virology.

In Chapter 8, I analysed how and why existing US dual-use oversight fails to capture the risks I identified in Part II. Existing DURC oversight is limited to specific experiments on a select list of agents, this list-based approach limits the scope of DURC but increases ease of implementation. A crucial limiting factor is the sole focus of DURC policies on microbiology research. Also experiments on non-pathogenic viral vectors create dual-use risks, and distinct non-microbiological dual-use risks arise from gene drive and neuroscience research. Additionally, I found that existing policies focus on *peak capabilities*, advances that create previously unattainable possibilities, and do not address dual-use risk from making existing methods more accessible. Lastly, I highlighted that the DURC policies only apply to a limited scope of organisations, research institutions funded by the United States federal government.

In Chapter 9, I examine how transferable insights can pose dual-use risks. This serves as an example of how dual-use risks can be created not just by high-risk experiments on dangerous pathogens but by a more diffuse set of research. I find that synthetic virology drives dual-use capabilities through advancing transferable and general-purpose viral engineering strategies that are applicable to pathogens. I introduced the term *transfer risks* to differentiate such transferable dual-use insights from research directly on pathogens, research which features *intrinsic risks*. In contrast to intrinsic risks, which are frequently conducted to understand infectious disease threats, research with transfer risks is frequently conducted for applications that do not necessitate creating dual-use knowledge. Therefore, other approaches for achieving the same benefits may be available which do not have dual-use potential. For instance, platform vaccines may be generated not only using viral methods but also non-viral methods, such as mRNA. Pursuing low-risk alternatives may thus be a promising strategy for reducing the risks of research with transfer risks. Additionally, biosecurity-by-design, including structured access, may be used to selectively disarm harmful applications of general-purpose technologies - examples are gene synthesis screening and the use of APIs for computational tools. My analysis of transfer risks illustrates how more diffuse advances generate dual-use risks. Thus, to prevent the misuse of synthetic viruses, governance approaches should not be limited to highest-risk research but need to consider dual-use risks across a broader spectrum.

In Chapter 10, I characterised the prevalent approach to dual-use management as *individual project oversight* and highlighted its limitations. Individual project oversight is characterised by risk-benefit assessment of individual research projects, which may then identify if risks outweigh benefits or if risks can be mitigated. I argued that while existing policies, such as P3CO, nominally involve consideration of low-risk alternatives, low-risk alternatives are considered with a narrow focus and need to meet a bar of providing the very same benefits. Existing individual project oversight could be strengthened by moving towards mandatory risk-benefit assessments conducted by an external independent agency. Even such

strong forms of individual project oversight would however be limited in their potential for risk reduction. Individual project governance does not comprehensively mitigate dual-use risks across a broader range of research, including research where risks are lower than benefits but still significant. A focus on individual projects features limited scope for encouraging consideration of benefits and risks in decisions between projects.

The last part of my analysis focused on approaches for governance that may be applicable to address the diffuse and diverse dual-use risks arising from synthetic virology. For inspiration, I look towards approaches used for the decarbonisation of the economy and mitigation of climate change. I discuss parallels between reducing carbon emissions and dual-use risks; both can be considered a negative externality of research and technology born by all of society. The need for the decarbonisation of the economy has meant that many disciplines, including economics and management, have considered interventions to shape the portfolio of energy technologies - and strategies have been trialled and tested in the real world.

In Chapter 11, I introduced *panoptic dual-use management*, an approach to dual-use risk mitigation which involves accounting for risks in decisions between projects and creating incentives for stakeholders to reduce dual-use risks. While individual project oversight cuts off a sliver of highest-risk research, I argued that panoptic dual-use management has the potential to shift the whole research portfolio towards the lower-risk end of the dual-use spectrum. I motivated the need for a panoptic dual-use management approach by analysing the carbon emission parallel and my findings on emerging dual-use risks. I argued that both constitute instances of the Tragedy of the Commons, a misalignment of incentives for individual stakeholders leading to an overall unfavourable outcome. I acknowledged differences between carbon emissions and dual-use risks, namely the disproportionate risks associated with the first disclosure of an

information hazard and Unilateralist's curse dynamics. Nevertheless, I argued that incentive systems similar to ones used for reducing carbon emissions could be beneficial for dual-use risk reduction. I also drew on the concept of *differential technology development* to inform panoptic dual-use management. Differential technology development highlights how changing the relative timing of research projects can have risk-reducing effects because of interactions with other technologies. It highlights that an important goal of dual-use management may involve delaying dual-use capabilities until the development of technologies for robust pandemic defence.

In Chapter 12, I then evaluated concrete ways to implement panoptic dual-use management. In doing so, I draw on lessons from climate change and leverage the theme of low-risk alternatives that I highlighted throughout my thesis. I propose to use incentive systems to preferentially induce the advancement of lower-risk and risk-reducing projects. I evaluate different ways in which incentives could be used to nudge different stakeholders to reduce dual-use research and advance risk-reducing approaches. I highlight the approaches that I identified as most interesting and promising in Box 14.1.

Box 14.1: Examples of panoptic dual-use management strategies

- Reducing likelihood of receiving funding for research proposals with greater dual-use potential, e.g. implemented through dual-use tiebreaker (see Chapter 13)
- Mandatory liability insurance for research with direct misuse risks
- Voluntary or mandatory public disclosure of aggregate data on dual-use research
- Tax benefits for companies engaging in risk-reducing practices
- Facilitating preferential advancement of low-risk solutions, for instance providing critical intellectual property for free
- Mandated use of screened DNA products for federally-funded research

In Chapter 13, I focus on a subset of panoptic dual-use management, *comparative risk-benefit assessment*. I note that in contrast to activities resulting in carbon emissions, governments control most of the research funding leading to dual-use capabilities. Thus, governmental funding bodies can directly curb dual-use risks. Funding bodies could compare the risks and benefits of research proposals and preferentially fund promising projects with less dual-use risk. I propose a concrete, actionable instantiation in the form of *dual-use tiebreakers*: for projects that funding bodies are uncertain whether to fund, dual-use risks could decide whether a project ends up with funding or not. Dual-use tiebreakers would directly shift what research is conducted and, importantly, also have incentive effects for researchers to propose high-promise proposals with less dual-use risk. To make dual-use tiebreakers actionable, I sketch out a tiered framework for scoring dual-use risks of research on viruses.

One overarching theme that emerges from Parts III and IV is that no silver bullet exists for dual-use management. Ultimately, with my proposals I hope to present new and updated ideas that may be adapted to become part of a diverse set of governance approaches.

14.4 A regime complex for mitigating dual-use risks

Others have previously pointed out that to manage dual-use risk in the life sciences, our best hope is to use and combine different approaches, to continuously experiment and learn (S. W. Evans et al. 2020; Palmer 2020). Again, one parallel is climate change. As mentioned in Chapter 10, climate change is not mitigated with a single governance regime; rather sustainable practices are advanced through a *regime complex* (Alter and Raustiala 2018). A regime complex emerges inadvertently when a range of actors try to address risks. Thus, a regime complex already exists to some degree for dual-use management. However, the current regime complex focuses on a narrow set of research in a narrow set of countries. As biotechnology progresses and misuse risks increase, a broader range of stakeholders need to consider where and how they

can contribute in their own context to this regime complex for dual-use management. Similar to mitigating climate change, the governance of dual-use risks needs to become a distributed and global undertaking.

14.4.1 Making dual-use management context-dependent

In previous chapters, I have presented different dual-use management approaches that may form building blocks for a global regime complex. I have described biosecurity-by-design and structured access, risk-benefit assessment and individual project oversight, and different panoptic dual-use management strategies. Panoptic dual-use management encompasses a diverse set of incentive-shaping interventions, from dual-use tiebreakers at funding bodies to social pressure for declaring dual-use research.

Each of these governance approaches may be finetuned across multiple dimensions; such finetuning may depend on who the relevant actor is and what projects they are trying to implement them for. Different actors - e.g. international organisations, governments, for-profit and non-profit institutions, or even individual researchers - may find that different panoptic dual-use management approaches work for them. In the following, I explore how dual-use management requires context-dependent fine-tuning based on (1) at what level governance is implemented and (2) the nature of target projects which defines the possibility of more or less flexible and binding governance measures. Afterwards, I then turn towards how differently context-adapted approaches can overlap to establish effective regime complexes.

First, each governance approach can be implemented at different levels, from international to national to local. For instance, local institutions or national agencies could conduct risk-benefit assessments of proposed research. The US has developed national guidance for gene synthesis screening, but also international organisations could develop such guidance. The WHO has created other guidelines, for instance, guidelines for mitigating laboratory accidents and dual-use management (World Health

Organisation 2020; 2022b). Binding regulations may be more difficult to achieve at an international level, but some precedent exists. For instance, the Montreal protocol successfully managed to globally coordinate the phase-out of ozone-depleting substances (McKenzie et al. 2019).

What governance approaches are best implemented at what level depends on what dual-use risks they aim to tackle. As discussed in Chapter 13, Steinbruner *et al.* proposed as part of their tiered governance framework that the highest concern research should be governed internationally. In contrast, research that is only potentially risky should be governed locally (Steinbruner et al. 2007). The worst cases of biological misuse affect the whole global community, thus universal risk mitigation practices should exist. At the same time, higher-level oversight reduces resolution and flexibility. Thus, higher-level oversight should be reserved for research with more defined and greater dual-use risks. Local approaches are more appropriate for research with potential concern, to allow flexible and tailored risk reduction.

A second important dimension is how governable target projects are and whether governance is implemented more or less formally. Implementation may be through legally binding regulations (hard law), voluntary guidelines and self-governance (soft law), or codes of conduct and education (informal measures). For instance, gene synthesis screening could be advanced under a guidance framework (soft law), as is currently the case in the US, or become a binding regulation (hard law). Comparative risk-benefit assessment could be implemented informally by educating scientists about the importance of considering dual-use risks when choosing research projects (informal measures) or could be implemented through formalised dual-use tiebreakers at funding bodies (soft law). I draw the categories of hard law, soft law, and informal measures from Jonathan Tucker's 2012 decision framework for dual-use management (Tucker 2012, 77). Tucker argues that whether hard law, soft law, or informal measures are applied should depend on the governability of a technology. For instance, faster advancing technologies may be better governed

by less formal measures that can be easily adapted to changing technologies. Tucker assesses governability not just based on rate of advancement, but also whether a technology is physical or non-physical and its level of maturity, its intersection of disciplines, and its international diffusion. I agree with Tucker that the governability of a technology depends on how diffuse it is and how fast it advances. Thus, DNA synthesis may be more governable than smaller-scale, specialised computational tools for viral vector engineering (see Figure 14.1b). Additionally, higher-risk research may warrant more binding regulations and greater associated costs to account for the unilateralist curse and the fact that the first disclosure of an information hazard is associated with the greatest risk.

Policymakers may also fine-tune dual-use management based on other dimensions. Different governance approaches may be more suitable for applying to the highest-risk research, others for applying to the whole spectrum of dual-use risks. Additionally, governance approaches may be targeted at different points across the research life cycle, depending on where governability and incentive effects are greatest. Lastly, how governance building blocks are assembled should be responsive to local context - for instance, the context of the country where they are considered. Governance approaches should fit existing governance structures and consider differences in biotechnological development and resources.

14.4.2 Context-dependent management assembles to effective regime complexes

In the following, I provide two examples of how different governance approaches could form context-dependent building blocks of a regime complex. The first example demonstrates how policymakers may combine governance approaches to cover a spectrum of dual-use risks (see Figure 14.1a). Well-circumscribed high-risk research could be subject to risk-benefit assessment to ensure benefits outweigh risks. For example, this includes PPP discovery and ePPP work (DUL5 in my scoring framework from Chapter 13). Risk-benefit assessment should be legally binding for such research and based on

international guidance. In conjunction, policymakers could use panoptic dual-use management approaches to incentivise risk reduction across a larger set of dual-use research. One example strategy would be a dual-use tiebreaker at governmental funding bodies, incentivising researchers to propose projects with less dual-use risk. Such a dual-use tiebreaker could apply to dual-use risks that are moderate in extent and relatively well-circumscribed. Lastly, policymakers might employ more local and informal governance approaches to capture less well-circumscribed or lower-risk research. Local dual-use experts could evaluate research risks across a wide range of research and help scientists identify risk-reducing measures, such as low-risk alternatives and biosecurity-by-design.

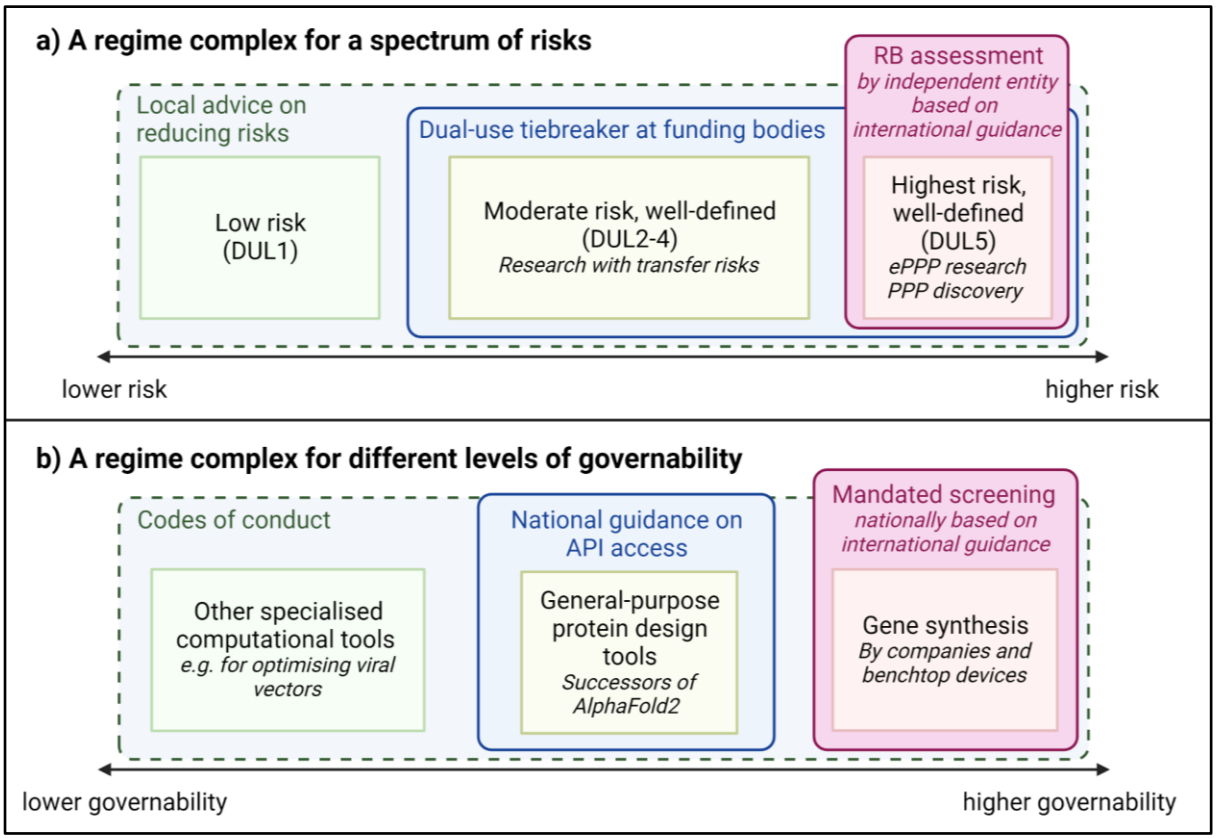


Figure 14.1: Regime complexes through combinations of governance approaches

a) Governance approaches may be combined to cover the spectrum of dual-use research. Higher-risk research warrants binding, individual project oversight based on international guidance. Moderate risk research warrants national incentive-based systems to reduce dual-use risks. Other research, including research with less well-defined risks, is captured by local identification and reduction of risks. b) Governance approaches may be combined to cover tools with different levels of governability

As a second example, I present how policymakers in governments and companies can use different governance approaches to address risks from general-purpose tools with different levels of governability. I focus on gene synthesis technology and computational tools for viral engineering. Gene synthesis involves physical products and is a relatively mature technology with relatively few advances. Thus, hard law governance measures, such as national requirements for gene synthesis screening, could be effective. Given the interconnectedness of the global market on gene synthesis, national requirements could be based on internationally consistent standards. Computational tools are software and are rapidly advancing, properties which make these tools less governable than gene synthesis. However, among computational tools, large-scale general-purpose tools, like cutting-edge large language models or general-purpose biodesign tools, are controlled by a small set of companies and thus may be more governable. In particular, future iterations of biodesign tools could require costly tailored training data and thus might only be developed by a small set of organisations. This small set of companies could lead the governance of these tools. Companies could employ structured access through APIs. Governmental policymakers could set national guidelines for standards of such APIs. Smaller-scale and more specialised models, including models regularly developed by academics for viral vector optimisation, feature low governability. Codes of conduct could cover such low governability tools and act as informal measures to encourage risk-reducing practices around data and code sharing.

Combinations of governance approaches, like the above examples, can increase the breadth and ambition of a regime complex for dual-use management. Not just governments but also civil society may contribute. As discussed for computational tools, company-led dual-use management will be crucial for fast moving technologies. Non-governmental organisations, similar to the Carbon Disclosure Project, could encourage infohazard sensitive disclosure of aggregated data on dual-use research. Forward-looking venture capital

firms could encourage biotechnology companies to develop biosecurity plans. Philanthropists may selectively facilitate the advancement of low-risk solutions. A broader, decentralised regime complex will facilitate dual-use management that is adaptive to new socio-technological developments (Chaffin, Gosnell, and Cosens 2014).

14.5 Overarching challenges

14.5.1 Uncertainty

Multiple overarching challenges limit advances in dual-use management. Addressing these challenges will require a sustained commitment. The first overarching challenge is uncertainty. Uncertainty about the likelihood and nature of misuse risks complicates decisions about what interventions or tradeoffs are justified. However, there are good arguments that the global community should take dual-use risks more seriously than is currently the case. First, biological events can be horrendous, killing many million humans. Deliberately released pathogens could be optimised for harm and have the potential to cause the collapse of society, an outcome that risks the continued existence of humanity and should be avoided at great priority. Thus, even a very small risk of catastrophic misuse should be sufficient to spark serious action. Furthermore, history demonstrates the existence of individuals with the intention to cause maximum harm, and the power and accessibility of biotechnology are increasing without any sign of slowing down. Even if it takes ten years for misuse risks to become significant, now is the time to set a path that considers whether and how to enter such a future. To reduce uncertainty about misuse, governments and other stakeholders could model the likelihood of misuse and the expected costs of dual-use research. Risk assessment models could draw on approaches developed in the financial industry and be informed by intelligence data on terror groups.

14.5.2 Apathy

The biggest challenge for trialling new dual-use management methods is apathy. Policymakers have limited attention and currently do not prioritise making progress on dual-use management. Apathy may be partially due to the rarity of past instances of misuse and uncertainty about the magnitude of risks, but also due to other factors.

Like pandemic preparedness more broadly, interventions to manage dual-use biotechnology feature a cycle of panic and neglect. The political will to tackle dual-use risks in the United States was high in the early 2000s - in the wake of 9/11 and the Anthrax attacks. At times, particularly newsworthy examples of dual-use research, such as the H5N1 enhancement experiments of 2011, revived this political will. However, more sustained attention by policymakers is required to trial new approaches and establish a broader regime complex for dual-use management. The current revisions of DURC and P3CO policies may provide a starting point.

Two additional factors may dissuade policymakers from spending their energy on advancing dual-use management. First, dual-use management is still a relatively niche concern and does not make for wins that appeal to voters. Second, and potentially more crucially, researchers have put up high resistance to the regulation of dual-use research. The scientific community has been set on self-governance - despite repeated evidence that self-governance is not sufficient (Salloch 2018). Even establishing risk-benefit assessment of high-risk ePPP research was difficult, and related debates continue (Duprex et al. 2015). If researchers do not believe dual-use risks warrant even a small amount of overhead, it will be difficult to advance conservative dual-use management approaches.

Apathy of dual-use management may be overcome by increasing awareness and education of dual-use risks and their mitigation. Also here, climate change can serve as an example: it took a long time for climate change to become a major topic on the international agenda. Climate activism started in the early 1970s, but it took until 1997 that countries formulated concrete goals for reducing carbon emissions as part of the Kyoto Protocol (Gupta 2010). Awareness of climate change was advanced through long years of campaigning, writing, and catastrophic weather events. For dual-use risks, the global community cannot rely on catastrophic events as a wake-up call, as such events might be associated with tremendous and irreversible harm. Now, in the wake of COVID-19 might be the ideal time to advance measures to manage dual-use risks relating to potential pandemic pathogens. Already the Fink report led with the recommendation to educate the scientific community about dual-use risks to allow scientists to fulfil their moral duty to prevent misuse (National Research Council 2004). As biotechnology becomes more powerful, risk reduction might become a core of scientific training.

14.5.3 Democracy

Governance systems should reflect the values of broader society. These values should guide how to trade off the benefits and costs of risk mitigation. Selgelid captures this in his principle "Good Governance - Democracy" (Selgelid 2016) (Appendix B). Thus, dual-use management requires ongoing societal input, input on dual-use management policies and input on specific research and technology projects. For instance, experiments like the recreation of horsepox should be conducted in consultation with the global public health and security community to ensure that it serves what wider society identifies as "good". Discussions may inform which research projects to pursue and how to shape research projects through biosecurity-by-design.

A large fraction of society is currently not even aware of the challenge of managing dual-use risks. In Chapter 2, I cite a survey in which 77% of participants were unaware that DURC is regularly conducted (MacIntyre et al. 2020). 64% deemed DURC unacceptable or were not sure of its acceptability - and interestingly, providing participants with more information on DURC decreased acceptance. A crucial step for advancing democratic input in dual-use management is educating society and establishing channels for societal input.

Societal input into science and technology cannot wait until the completion of a project - at that stage, input comes too late. Responsible innovation efforts already aim at establishing channels for societal input. An example are the consensus conferences hosted by the Danish Board of Technology (Dryzek and Tucker 2008). However, scientists must also take up leadership in seeking public input. Kevin Esvelt has previously highlighted how the development of gene drives may drive science to become open and responsive (Esvelt 2016). Seeking public input does not have to involve hosting town hall meetings, as Esvelt has done on Nantucket, a potential site of gene drive experiments (Specter 2016). Esvelt proposes a lower bar for starting: openly publishing grant proposals and pre-registering experiments before the initiation of research (Esvelt 2016; 2017).

The timescale over which work proceeds defines the possibility of societal input. If technology development is dyssynchronous (i.e. faster) than the rate at which broader society comes to understand and think through implications, then there is an issue. The accelerating pace of biotechnology democratisation contributes to the challenge of dyssynchronous timescales (S. S. Jackson et al. 2019). Scientists and technology developers will sometimes have to halt and allow society to catch up. Governance

systems may also establish speed bumps to align timescales.[61] As Sam Weiss Evans argues, the relationship between science and society might have to change fundamentally (S. W. Evans 2022). Science must not operate separately from society but become a part of it.

New consensus-finding tools may help foster democratic input into dual-use management. One example is the platform Polis, which I mentioned as a tool to help create a tiered dual-use scoring framework in Chapter 13 ('Polis' 2022). Policymakers, funding bodies, and academics could use such tools for crowdsourcing input on how to trade off risks.

14.5.4 Global coordination and equity

The last overarching challenge is the global nature of dual-use risks. Because of the scale of biological risks, dual-use risks of synthetic virology impact all of humanity. No nation is safe from catastrophic pandemics until the world is safe. Thus, the ethical acceptability of dual-use research depends not only on local values but also on international ones. Selgelid recognises this as part of the principle "International Outlook and Engagement" (Selgelid 2016) (see Appendix II).

Global coordination on dual-use management is extremely challenging. Enforceable international rules are presently unattainable. Similar to the mitigation of climate change, national interests are deeply interwoven with risk mitigation. As countries struggle to dominate the global bioeconomy, they may neglect risk mitigation. The United States currently has the leading basic science research infrastructure, but China and other players like India are catching up (R. Carlson and Wehbring 2020). Thus, dual-use

[61] An idea that David Relman has mentioned on multiple occasions, including relating to the risks of publishing step-by-step protocols for creating potential pandemic pathogens.

research will likely diffuse and international coordination of governance will become increasingly important.

The difficulty of global coordination is exacerbated by economic disparity. Countries with limited resources might struggle to prioritise dual-use management. Worrying about abstract, potential future dual-use risks might be seen as a privilege of wealthy countries. Developing countries may see security and equity as clashing concepts. For instance, countries of the Global South have previously challenged security-motivated export controls of the Australia Group under Article X of the BWC, the article which demands international cooperation and the exchange of biological agents for peaceful use. However, equity and security are not necessarily always opposed. As I argued in Chapter 4, APIs for computational tools may both establish security-sensitive structured access and reduce requirements for technical skills and computational resources.

When considering dual-use management in light of economic disparity, climate change may again serve again as a parallel. Developed countries are said to have accrued a "climate debt" because they disproportionately contributed to climate change, which now disproportionately threatens developing countries (Pickering and Barry 2012). While biological events also disproportionately threaten countries with less developed healthcare systems, the historical debt around dual-use risks is smaller than that for carbon emissions. Scientific advances are more likely to benefit all of humanity than carbon emissions, and the greatest dual-use risks have likely not yet materialised. Nevertheless, one might identify a "dual-use debt" in the form of a dual-use skill gap. To date, developed countries have trained the most individuals with dual-use skills and have the most dual-use laboratory facilities. Suppose other countries would train similar levels of individuals in dual-use skills: in that case, the number of individuals capable of engineering viruses would multiply - as would risk of misuse. As biotechnology becomes more democratised and more

globally distributed, all countries need to adopt practices to mitigate dual-use risks, including during the training of new scientists - and wealthy countries should support others in this process. Some of this work is already on the way. For instance, Africa CDC has created a biosafety and biosecurity strategy and is becoming a global biosecurity leader with the support of the G7 (Africa CDC 2021; Global Partnership Against the Spread of Weapons and Material of Mass Destruction 2022).

Regime complexes may help to tackle global coordination and equity challenges. Universal governance approaches, such as the ones I have outlined in this thesis, may be adapted based on local considerations for use in different settings. Furthermore, even without binding international agreements, the international regime complex for mitigating climate change is driving global decarbonisation. Thus, loose collaborations and agreements may also be sufficient to drive an effective global regime complex for dual-use management.

14.6 Avenues for further research

This thesis highlights many opportunities for further research and testing of whether the introduced ideas can survive confrontation with the real world. Policymaking is an iterative process; proposing interventions from the academic ivory tower only goes so far. To further advance the concept of panoptic dual-use management and concrete proposals such as dual-use tiebreakers, these ideas would need to be socialised more widely to stakeholders and refined based on their input. Concrete exploration is needed on how to turn concepts like dual-use tiebreakers into implementation-ready policy - including for non-US settings.

Next to putting identified ideas to a test, there is also a need for a larger set of follow-on conceptual, qualitative, and technical research. First, I highlighted the important intersection between biotechnology

and artificial intelligence in Chapter 7. How to govern these fast-evolving dual-use capabilities is a research project of high urgency. Second, there is clear potential to dig more deeply into incentive systems for different stakeholders and thinking for example about how concretely academic researchers could be incentivised to consider dual-use risks in their decisions, and how dual-use risks could be considered as part of crucial criteria for academic career advancement. Third, investigating how the identified trends apply to the Global South would be interesting and important; an analysis of what dual-use research is conducted in less well-resourced settings could shed light on the most urgent and promising avenues for reducing dual-use risks in a broader global context. Part of such research could include examining the dual-use skill gap and security implications of its closure.

14.7 The big picture view: future-proof pandemic prevention

To end this thesis, I want to put the role of dual-use management into a broader perspective of pandemic prevention. As discussed in Chapter 10, dual-use management may be seen as a set of interventions to prevent misuse until society has developed robust defences. Dual-use management tries to constrain dual-use capabilities to prevent misuse. However, this feat won't be possible for all actors. Dual-use management can, at best, prevent misuse by lone wolfs and less well-resourced groups, but it can only hope to raise the bar for well-resourced and sophisticated actors.

Dual-use management can be complemented with other strategies to reduce misuse risks broadly. To prevent well-resourced actors, such as states, from the offensive use of synthetic viruses, efforts to shape their intent are crucial. Such efforts may include advancing a new compliance mechanism for the BWC or achieving deterrence by denial, making biological weapons unattractive through the advancement of defensive technologies. To catch more sophisticated non-state actors, improving intelligence and law enforcement capabilities might be key. Any project involving design-build-test cycles to enhance a

pathogen will likely require longer-term and repeated access to synthetic DNA, laboratory facilities, test subjects, and human resources. Creating intelligence systems that pick up on relevant signatures could help identify potential perpetrators before they can achieve their goals. Additionally, better mental health provisions for researchers with relevant molecular biology skills could also contribute to averting instances of misuse and provide channels for flagging concerns.

Efforts to prevent misuse must be coupled with sustained investments in defensive technologies and the strengthening of health care resilience. Such investments will hopefully eventually make the defence against pandemics a viable option and thus will help close the risk window for the catastrophic misuse of biotechnology. Pandemic preparedness plans of White House and G7 prioritise the advancement of vaccine platform technologies, including the creation of prototype vaccines across viral families (The White House 2021; Pandemic Preparedness Partnership 2021). The 100 Day Mission aims to significantly shorten vaccine development. Fast vaccine development can be complemented with rapid sequencing-based detection and needs to be coupled to advances in vaccine manufacturing and distribution. However, the global community needs to consider a response that does not rely on total immunisation. It is unclear whether immunisation will be safe and effective against any pathogen.

In the face of a catastrophic pandemic with a high fatality rate, next-generation personal protective equipment could be crucial to protect essential workers who keep society running. Together with rapid diagnostic tests and new methods for contact tracing, better personal protective equipment might eventually halt the majority of pandemics. Advances in indoor air filtration may also contribute. Far-UVC, light with wavelengths between 200-230 nm, may inactivate pathogens while being safe for human skin and eyes (Blatchley et al. 2022). If efficacy and safety are confirmed and emitters for far-UVC are improved, large-scale deployment of far-UVC might significantly reduce indoor transmission.

Technology to respond to pandemics is improving steadily. The development of a vaccine for COVID-19 within a single year was an unprecedented feat. Humanity will eventually be able to stop an emerging pathogen in its tracks. However, even if we try everything to speed up defensive technologies, achieving a point of robust defence will likely take decades - decades during which dual-use research may advance offensive capabilities. Dual-use management is our best bet at delaying offensive capabilities and preventing catastrophic misuse. I hope my ideas will contribute to this vital mission.

GLOSSARY

Term	Definition
Application programming interface (API)	The software by which users communicate with applications like tools and databases, and applications connect with each other.
Barcoding	Short sequences of DNA inserted into viruses allow identification of individual viral mutants after selection. This enables parallel directed evolution of large numbers of viruses in a single experiment.
Biodesign tools	Artificial intelligence models that are trained on biological data that can help design new proteins or other biological agents.
Biosecurity-by-design	Designing biotechnology in a way that prevents or reduces risk of misuse. The concept was first developed for a 2016 National Defense University meeting and since expanded upon by the biosecurity community (DiEuliis and Lutes 2016; Budeski 2018).
Comparative risk-benefit assessment	Consideration of benefits and risks in decisions between research projects.
CRISPR/Cas9	A versatile and accurate approach for cutting DNA, in which the bacterial CRISPR-associated protein 9 (Cas9) is guided by a CRISPR (clustered regularly interspaced short palindromic repeats) sequence to the target site for cleavage. This mechanism was discovered in bacteria, where it is used as a form of immune system against unwanted genomic insertions.
Differential technology development	A principle to leverage risk-reducing interactions between technologies by affecting their relative timing. (Sandbrink, Hobbs, et al. 2022)
DURC	Dual Use Research of Concern (DURC) is life sciences research that, based on current understanding, can be reasonably anticipated to provide knowledge, information, products, or technologies that could be directly misapplied to pose a significant threat with broad potential consequences to public health and safety, agricultural crops and other plants, animals, the environment, materiel, or national security.[62]
Dual-use research (DUR)	Research with legitimate applications that features potential to be misapplied for causing harm.
Gain-of-function (GOF) research	Broadly, any research trying to enhance a property of an organism. Here, this term refers to research involving the enhancement of potential pandemic pathogens for human transmissibility and lethality.

[62] U.S. Department of Health and Human Services. United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern. 2012.

Gene drive	A genetic engineering approach, for instance based on CRISPR/Cas9, which propagates a particular (modified) gene throughout a population. Gene drives have been explored for reducing fertility or disease-carrying capacity of infectious disease vectors.
Gene/DNA synthesis screening	Gene synthesis products (such as DNA) that have been checked to see if they encode for regulated pathogens or other dangerous proteins. Gene synthesis companies that screen their orders in this way frequently also screen customers for their institutional affiliation.
Gene therapy	Medication to address genetic disorders, such as the absence of certain genes that results in deficiencies in their gene products. Usually, an artificial copy of the gene is inserted through a delivery device such as a viral vector.
Governance	"The norms, values and rules of the processes through which public affairs are managed so as to ensure transparency, participation, inclusivity and responsiveness. Governance also represents the structures and processes that are designed to ensure accountability, transparency, responsiveness, adherence to the rule of law, stability, equity and inclusiveness, empowerment, and broad-based participation."[63]
Immunotherapy	Medications that induce or modulate immune responses, for instance for fighting cancer.
Individual project oversight	Risk-benefit assessments or risk mitigation strategies that focus on an individual project, experiment, or technology.
Information hazard	"A risk that arises from the dissemination or the potential dissemination of (true) information that may cause harm or enable some agent to cause harm" (Bostrom 2012)
Intrinsic risk	The risk of misuse of knowledge, methods, or technologies developed on or resulting in (high risk) agent A, where as a result agent A poses a significant threat with broad potential consequences.
Large language model (LLM)	An artificial intelligence model trained on a large amount of text, which is very good at generating answers to prompts. These models can be deployed as chatbots, such as chatGPT, or can be used as the basis for agents which can automatically interface with more specialised tools (e.g. ChemCrow).
Structured access	Access to certain technologies through security-sensitive methods, which prevent misuse. This concept was initially introduced for the governance of artificial intelligence systems. (Shevlane 2022)
Non-state actor	Actors that are not states, such as individuals or groups. In the biosecurity

[63] Definition used in 2022 WHO Global Guidance Framework for the Responsible Use of the Life Sciences, which in turn adopted this definition from the 2021 WHO report "Human genome editing: a framework for governance"

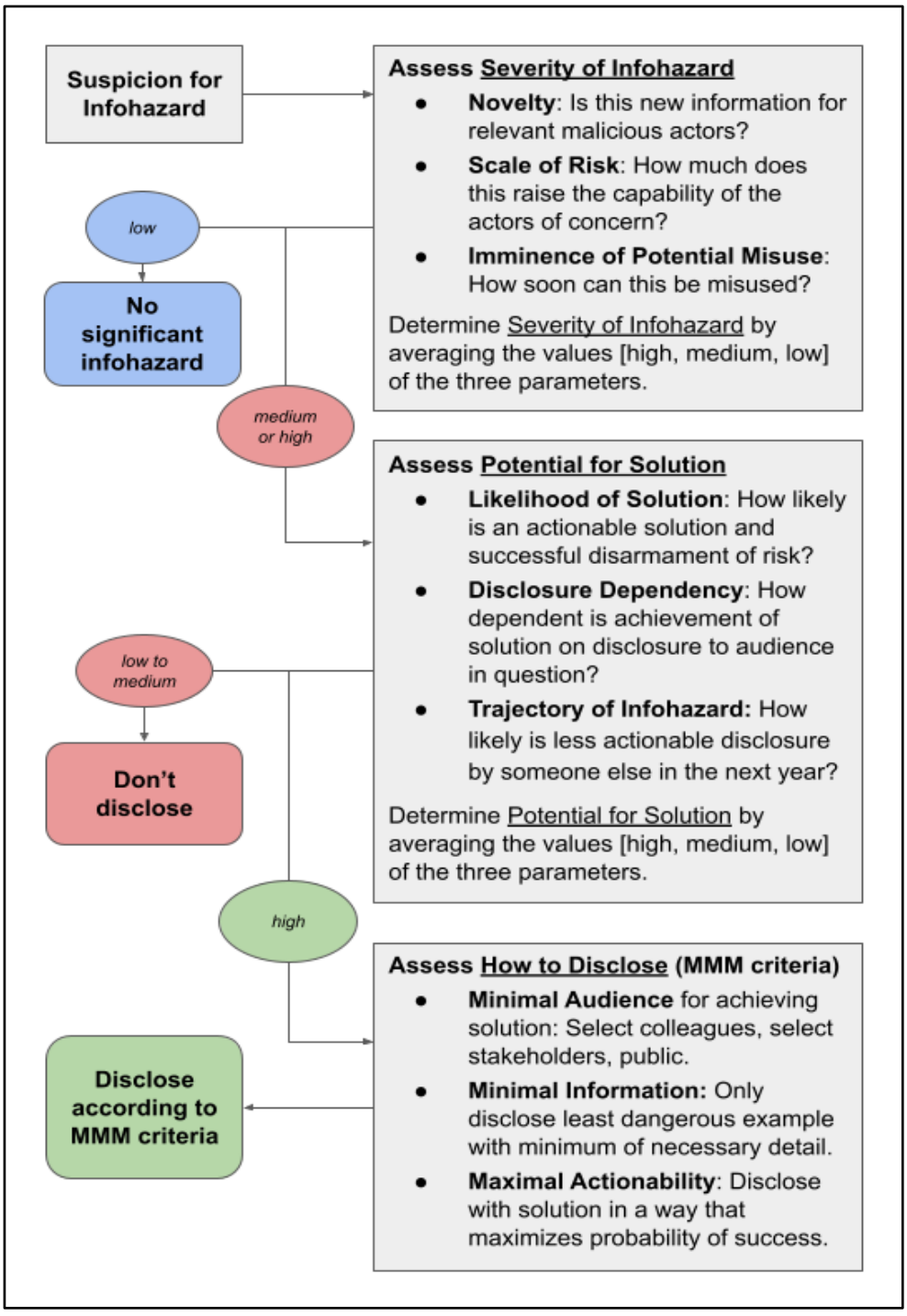
	context often used to describe perpetrators of biological misuse, such as ideologically or otherwise motivated terrorist groups or individuals.
Open Science	The movement to make science more accessible, transparent, and reproducible.
Pandemic-capable pathogen	A pathogen that if introduced in the human population is with significant confidence capable of causing a pandemic. A virus may be classified as pandemic-capable if it features similar binding affinity to human receptors as comparable endemic viruses and no preexisting immunity in humans that would limit its spread.
Panoptic dual-use management	Reducing dual-use risks through accounting for risks in decisions between projects and creating incentives for stakeholders to reduce dual-use risks.
Peak capabilities	Capabilities that enable people to do things (with dual-use potential) that no one could do before.
Polymerase chain reaction (PCR)	A popular molecular biology method to make many copies of a specific section of DNA, which is selected using a so-called primer complementary to the sequence in question. PCR machines are available in almost all molecular biology laboratories and work by cycling through different temperatures which repeatedly activate enzymes involved in the copying of DNA.
Potential pandemic pathogen (PPP)	Pathogens that are likely to be highly transmissible and capable of causing significant harm in humans.[64]
Regime complex	A combination of partially overlapping governance measures adds up to form a regime complex. Despite not being centralised or globally coordinated, regime complexes can be effective in addressing difficult governance challenges and allow for consideration of local contexts. The regime complex framing was first introduced for climate change mitigation. (Alter and Raustiala 2018)
Reverse genetics platforms	Non-virus specific approaches for turning synthetic DNA into a functional virus. These may replace helper virus-dependent methods for viral booting that require more specialist skills and resources.
Tacit knowledge	All knowledge that is difficult to convey through language, such as the ability to ride a bike. In this context, it refers to knowledge for laboratory techniques, experimental protocols, pathogen specific work acquired through training and experience.
Transfer risk	The risk of misuse of knowledge, methods, or technologies developed on (low risk) agent A which can be applied to an agent B, where B as a result poses a significant threat with broad potential consequences.

[64] Adapted from P3CO policy definition. U.S. Department of Health and Human Services. Framework for Guiding Funding Decisions about Proposed Research Involving Enhanced Potential Pandemic Pathogens. 2017 https://www.phe.gov/s3/dualuse/Documents/P3CO.pdf (accessed 16 Nov2020).

Transmissible vaccines	Vaccines capable of autonomously spreading through wild animal reservoirs for proactively addressing spillover of viruses into humans. Viral vectors engineered for transmissibility, genetic stability, and evasion of pre-existing immunity are one proposed method.
Viral booting	The synthesis of infectious viruses from synthetic DNA.

APPENDIX A: Framework for decisions on information hazards

When deciding whether to disclose a concrete piece of dual-use information, I considered the following framework (of my own creation) for assessing the magnitude of its information hazard potential, the benefits of disclosure, and how to disclose it. Thank you to Kevin Esvelt, Ethan Alley, and Will Bradshaw for advice on designing this framework.



APPENDIX B: Ethical principles for decision- and policy-making on dual-use research

Adapted from principles for decision- and policy-making on gain of function research in Selgelid 2016.

DUR = dual-use research.

Ethical Principle	Explanation
Research Imperative	The ethical acceptability of DUR posing extraordinary risks partly depends on the importance of the research question it aims to address.
Proportionality	The ethical acceptability of extraordinarily risky DUR partly depends on the extent to which there is reasonable expectation that the research in question will (1) yield answers to the target question and (2) ultimately result in benefits that outweigh risks involved.
Minimisation of Risks	Other things being equal, the ethical acceptability of DUR is a function of the degree to which (1) there is confidence that no less risky forms of research would be equally beneficial and (2) reasonable steps have been made to minimise risks of the DUR in question.
Manageability of Risks	Other things being equal, the more manageable the risks of DUR, the more ethically acceptable the study or technology would be. Conversely, the more important/beneficial DUR is expected to be, the more we should be willing to accept potentially unmanageable risks.
Justice	Because justice requires fair sharing of benefits and burdens, the ethical acceptability of DUR depends on the degree to which (1) risks fall on some people more than others, (2) risks fall on those who are unlikely to benefit, and/or (3) any resulting harms are uncompensated.
Good Governance — Democracy	DUR decision- and policy-making should (insofar as possible) reflect the ultimate values, value weightings, and risk-taking strategies of public citizens.
Evidence	Decision- and policy-making regarding DUR should be based on more/better evidence regarding risks, benefits, (means of) risk minimisation, who is likely to benefit or be harmed by research, and the values, value weightings, and risk-taking strategies of public citizens.
International Outlook and Engagement	Because risks and benefits of DUR (can) affect the global community at large, the ethical acceptability of DUR partly depends on the extent to which it is accepted internationally. Decision- and policy-making regarding DUR should (insofar as possible) involve consultation, negotiation, coordination, and related forms of active engagement with other countries.

APPENDIX C: Detailed statement of authorship

This thesis presents some co-authored material, which I recognise in this appendix.

Chapter 4

Chapter 4 presents content from the following cop-authored preprint [1]. This material has been completely rewritten and paraphrased for the context of this thesis. Figure 4.1 has been slightly adapted from the pre-print (which in turn adapted this figure from Carlson et al. 2021).

[1] Sandbrink, Jonas B., Janvi Ahuja, Jacob Swett, Gregory Koblentz, and Claire Standley. 2022. "Mitigating Biosecurity Challenges of Wildlife Virus Discovery and Characterisation." SSRN Scholarly Paper ID 4035760. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4035760.

I conceptualised this preprint and wrote its original draft. Co-authors helped improve this work and edited the draft, including by providing technical input and additional references. The authorship statement of [1] reads: "JBS: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. JA, JLS, GDK, CJS: Writing - review & editing."

Chapter 5

Chapter 5 presents content from papers [2] and [3]. This material has been rewritten and paraphrased for the context of this thesis. Figure 5.1 was reproduced from [2].

[2] Sandbrink, Jonas B., and Gregory D. Koblentz. 2022. "Biosecurity Risks Associated with Vaccine Platform Technologies." *Vaccine* 40 (17): 2514–23. https://doi.org/10.1016/j.vaccine.2021.02.023.

I conceptualised this paper and wrote its original draft. My co-author helped improve this work and edited the draft, including by providing his expert input and additional references. The authorship statement of [2] reads: "Jonas B. Sandbrink: Conceptualization, Investigation, Writing - original draft. Gregory D. Koblentz: Writing - review & editing."

[3] Sandbrink, Jonas B., Matthew C. Watson, Andrew M. Hebbeler, and Kevin M. Esvelt. 2021. "Safety and Security Concerns Regarding Transmissible Vaccines." *Nature Ecology & Evolution* 5 (4): 405–6. https://doi.org/10.1038/s41559-021-01394-3.

I wrote the original draft of this correspondence piece, with substantial input by Kevin Esvelt. Matthew Watson flagged the paper and suggested to consider writing a correspondence piece. All co-authors performed edits on my original draft. There is no official authorship statement, because of the correspondence format of this piece.

Chapter 6

Chapter 6 presents content from paper [4]. This material has to been rewritten and paraphrased for the context of this thesis. Figure 6.1 was reproduced from [4].

[4] Sandbrink, Jonas B., Ethan C. Alley, Matthew C. Watson, Gregory D. Koblentz, and Kevin M. Esvelt. 2022. "Insidious Insights: Implications of Viral Vector Engineering for Pathogen Enhancement." *Gene Therapy*, March. https://doi.org/10.1038/s41434-021-00312-3.

I conceptualised this paper and wrote its original draft. Co-authors helped improve this work and edited the draft, including by providing technical input and additional references. The authorship statement of [4] reads: "JBS: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. JA, JLS, GDK, CJS: Writing - review & editing."

Chapter 7

Chapter 7 presents the content of a single-authored preprint [5] . Chapter 7 is based on an earlier version of preprint [5], and thus features a large number of sections with word-for-word overlap with the preprint. Figure 7.1 was reproduced from [5].

[5] Sandbrink, Jonas B. 2023. "Artificial intelligence and biological misuse: Differentiating risks of language models and biodesign tools." arXiv preprint. https://arxiv.org/abs/2306.13952.

Chapter 11

Chapter 11 section 11.2 presents content of the co-authored preprint [6]. This material has been completely rewritten and paraphrased for the context of this thesis. Figure 11.2 was slightly adapted from [6].

[6] Sandbrink, Jonas B., Hamish Hobbs, Jacob Swett, Allan Dafoe, and Anders Sandberg. 2022. "Differential Technology Development: A Responsible Innovation Principle for Navigating Technology Risks." SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=4213670.

This paper is based on a concept originally introduced by Nick Bostrom, and fleshed out by a variety of thinkers over the years. Hamish Hobbs had written a first draft of this paper, which I ended up rewriting nearly completely. Many examples and conceptual details are my work, however many components were also taken from Hamish Hobbs' previous draft and the contributions of the other co-authors. The authorship statement reads: "Jonas B. Sandbrink: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. Hamish Hobbs: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. Jacob L. Swett: Conceptualization, Writing - review & editing. Allan Dafoe: Conceptualization, Writing - review & editing. Anders Sandberg: Conceptualization, Writing - review & editing."

REFERENCES

Adalja, Amesh A, Matthew Watson, Anita Cicero, and Thomas V. Inglesby. 2019. 'Vaccine Platforms: State of the Field and Looming Challenges'. Johns Hopkins Center for Health Security. https://www.centerforhealthsecurity.org/our-work/pubs_archive/pubs-pdfs/2019/190423-OPP-platform-report.pdf.

Administration for Strategic Preparedness & Response. 2020. 'Comments Received in Response to Federal Register Notice 2020-18444, Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA'.

Adriouch, Sahil, Emilie Franck, Laurent Drouot, Carole Bonneau, Nelly Jolinon, Anna Salvetti, and Olivier Boyer. 2011. 'Improved Immunological Tolerance Following Combination Therapy with CTLA-4/Ig and AAV-Mediated PD-L1/2 Muscle Gene Transfer'. *Frontiers in Microbiology* 2: 199. https://doi.org/10.3389/fmicb.2011.00199.

Africa CDC. 2020. 'African Researchers Lead Scientific Coalition Developing Surveillance System for Detecting Emerging Pandemics in Real-Time'. *Africa CDC* (blog). 2020. https://africacdc.org/news-item/african-researchers-lead-scientific-coalition-developing-surveillance-system-for-detecting-emerging-pandemics-in-real-time/.

———. 2021. 'Biosafety and Biosecurity Initiative 2021 - 2025 Strategic Plan'. https://africacdc.org/download/biosafety-and-biosecurity-initiative-2021-2025-strategic-plan/.

Ahn, K., T. H. Meyer, S. Uebel, P. Sempé, H. Djaballah, Y. Yang, P. A. Peterson, K. Früh, and R. Tampé. 1996. 'Molecular Mechanism and Species Specificity of TAP Inhibition by Herpes Simplex Virus ICP47.' *The EMBO Journal* 15 (13): 3247–55. https://doi.org/10.1002/j.1460-2075.1996.tb00689.x.

Ahn, Kwangseog, Albrecht Gruhler, Begona Galocha, Thomas R. Jones, Emmanuel J. H. J. Wiertz, Hidde L. Ploegh, Per A. Peterson, Young Yang, and Klaus Früh. 1997. 'The ER-Luminal Domain of the HCMV Glycoprotein US6 Inhibits Peptide Translocation by TAP'. *Immunity* 6 (5): 613–21. https://doi.org/10.1016/S1074-7613(00)80349-0.

Alizon, S., A. Hurford, N. Mideo, and M. Van Baalen. 2009. 'Virulence Evolution and the Trade-off Hypothesis: History, Current State of Affairs and the Future'. *Journal of Evolutionary Biology* 22 (2): 245–59. https://doi.org/10.1111/j.1420-9101.2008.01658.x.

Alter, Karen J., and Kal Raustiala. 2018. 'The Rise of International Regime Complexity'. *Annual Review of Law and Social Science* 14 (1): 329–49. https://doi.org/10.1146/annurev-lawsocsci-101317-030830.

Amarilla, Alberto A., Julian D. J. Sng, Rhys Parry, Joshua M. Deerain, James R. Potter, Yin Xiang Setoh, Daniel J. Rawle, et al. 2021. 'A Versatile Reverse Genetics Platform for SARS-CoV-2 and Other

Positive-Strand RNA Viruses'. *Nature Communications* 12 (1): 3431. https://doi.org/10.1038/s41467-021-23779-5.

Amman, Brian R., Brian H. Bird, Ibrahim A. Bakarr, James Bangura, Amy J. Schuh, Jonathan Johnny, Tara K. Sealy, et al. 2020. 'Isolation of Angola-like Marburg Virus from Egyptian Rousette Bats from West Africa'. *Nature Communications* 11 (1): 510. https://doi.org/10.1038/s41467-020-14327-8.

Amodei, Dario. 2023. 'Written Testimony of Dario Amodei, Ph.D.' US Senate, Judiciary Committee Subcommittee on Privacy, Technology, and the Law. https://www.judiciary.senate.gov/imo/media/doc/2023-07-26_-_testimony_-_amodei.pdf.

Anderljung, Markus, and Julian Hazell. 2023. 'Protecting Society from AI Misuse: When Are Restrictions on Capabilities Warranted?' arXiv. https://doi.org/10.48550/arXiv.2303.09377.

Anderson, Jeffrey J., David Rode, Haibo Zhai, and Paul Fischbeck. 2021. 'A Techno-Economic Assessment of Carbon-Sequestration Tax Incentives in the U.S. Power Sector'. *International Journal of Greenhouse Gas Control* 111 (October): 103450. https://doi.org/10.1016/j.ijggc.2021.103450.

Armendariz, A, P D'haeseleer, D Gillum, D Grushkin, E Harness, T. Kuiken, and J Molloy. 2020. *Full Community Biology Biosafety Handbook*. https://docs.google.com/document/d/1Qkc2uCAcLX45b0GjSGZohweelJ-vDOhX5MDSf6F4MEI/edit?usp=embed_facebook.

Arrow, Kenneth. 1970. 'Political and Economic Evaluation of Social Effects and Externalities'. In *The Analysis of Public Output*, 1–30. NBER. https://www.nber.org/books-and-chapters/analysis-public-output/political-and-economic-evaluation-social-effects-and-externalities.

Atlas, Ronald M., and Malcolm Dando. 2006. 'The Dual-Use Dilemma for the Life Sciences: Perspectives, Conundrums, and Global Solutions'. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 4 (3): 276–86. https://doi.org/10.1089/bsp.2006.4.276.

Baek, Minkyung, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, et al. 2021. 'Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network'. *Science* 373 (6557): 871–76. https://doi.org/10.1126/science.abj8754.

Bai, Yuntao, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, et al. 2022. 'Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback'. arXiv. https://doi.org/10.48550/arXiv.2204.05862.

Bakker, Kevin M., Tonie E. Rocke, Jorge E. Osorio, Rachel C. Abbott, Carlos Tello, Jorge E. Carrera, William Valderrama, Carlos Shiva, Nestor Falcon, and Daniel G. Streicker. 2019. 'Fluorescent Biomarkers Demonstrate Prospects for Spreadable Vaccines to Control Disease Transmission in

Wild Bats'. *Nature Ecology & Evolution* 3 (12): 1697–1704. https://doi.org/10.1038/s41559-019-1032-x.

Barash, Jason R., and Stephen S. Arnon. 2014. 'A Novel Strain of Clostridium Botulinum That Produces Type B and Type H Botulinum Toxins'. *The Journal of Infectious Diseases* 209 (2): 183–91. https://doi.org/10.1093/infdis/jit449.

Barker, Stephanie. 2020. 'Covid Vaccine Front-Runner Is Months Ahead of Her Competition'. *Bloomberg*, 15 July 2020. https://www.bloomberg.com/news/features/2020-07-15/oxford-s-covid-19-vaccine-is-the-coronavirus-front-runner.

Bartoszewicz, Jakub M, Anja Seidel, and Bernhard Y Renard. 2021. 'Interpretable Detection of Novel Human Viruses from Genome Sequencing Data'. *NAR Genomics and Bioinformatics* 3 (1). https://doi.org/10.1093/nargab/lqab004.

BBSRC, MRC, and Wellcome. 2015. 'BBSRC, MRC and Wellcome Trust Position Statement on Dual Use Research of Concern and Research Misuse'. https://cms.wellcome.org/sites/default/files/wtp059491.pdf.

Benner, Steven A., and A. Michael Sismour. 2005. 'Synthetic Biology'. *Nature Reviews Genetics* 6 (7): 533–43. https://doi.org/10.1038/nrg1637.

Berg, P., D. Baltimore, S. Brenner, R. O. Roblin, and M. F. Singer. 1975. 'Summary Statement of the Asilomar Conference on Recombinant DNA Molecules'. *Proceedings of the National Academy of Sciences of the United States of America* 72 (6): 1981–84. https://doi.org/10.1073/pnas.72.6.1981.

Berkovitz, Dan M. 1989. 'Price-Anderson Act: Model Compensation Legislation- The Sixty-Three Million Dollar Question'. *Harvard Environmental Law Review* 13 (1): 1–68.

Blatchley, Ernest R., David J. Brenner, Holger Claus, Troy E. Cowan, Karl G. Linden, Yijing Liu, Ted Mao, et al. 2022. 'Far UV-C Radiation: An Emerging Tool for Pandemic Control'. *Critical Reviews in Environmental Science and Technology* 0 (0): 1–21. https://doi.org/10.1080/10643389.2022.2084315.

Boddie, Crystal, Matthew Watson, Gary Ackerman, and Gigi Kwik Gronvall. 2015. 'Assessing the Bioweapons Threat'. *Science* 349 (6250): 792–93. https://doi.org/10.1126/science.aab0713.

Boiko, Daniil A., Robert MacKnight, and Gabe Gomes. 2023. 'Emergent Autonomous Scientific Research Capabilities of Large Language Models'. arXiv. https://doi.org/10.48550/arXiv.2304.05332.

Bosaeed, Mohammad, Hanan H. Balkhy, Sultan Almaziad, Haya A. Aljami, Hind Alhatmi, Hala Alanazi, Mashael Alahmadi, et al. 2022. 'Safety and Immunogenicity of ChAdOx1 MERS Vaccine Candidate in Healthy Middle Eastern Adults (MERS002): An Open-Label, Non-

Randomised, Dose-Escalation, Phase 1b Trial'. *The Lancet Microbe* 3 (1): e11–20. https://doi.org/10.1016/S2666-5247(21)00193-2.

Bostrom, Nick. 2012. 'Information Hazards: A Typology of Potential Harms from Knowledge', March.

———. 2014. *Superintelligence: Paths, Dangers, Strategies*. Illustrated edition. Oxford: OUP Oxford.

Bostrom, Nick, Thomas Douglas, and Anders Sandberg. 2016. 'The Unilateralist's Curse and the Case for a Principle of Conformity'. *Social Epistemology* 30 (4): 350–71. https://doi.org/10.1080/02691728.2015.1108373.

Bou-Antoun, Sabine, Ceire Costelloe, Kate Honeyford, Mahsa Mazidi, Benedict W J Hayhoe, Alison Holmes, Alan P Johnson, and Paul Aylin. 2018. 'Age-Related Decline in Antibiotic Prescribing for Uncomplicated Respiratory Tract Infections in Primary Care in England Following the Introduction of a National Financial Incentive (the Quality Premium) for Health Commissioners to Reduce Use of Antibiotics in the Community: An Interrupted Time Series Analysis'. *Journal of Antimicrobial Chemotherapy* 73 (10): 2883–92. https://doi.org/10.1093/jac/dky237.

Boutin, Laetitia, Estelle Mosca, and Frédéric Iseni. 2022. 'Efficient Method for Generating Point Mutations in the Vaccinia Virus Genome Using CRISPR/Cas9'. *Viruses* 14 (7): 1559. https://doi.org/10.3390/v14071559.

Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. New York, NY: Oxford University Press.

Bruggen, Koos van der. 2013. 'Biosecurity and the Just-War Tradition'. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 207–22. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.18.

Bryant, Drew H., Ali Bashir, Sam Sinai, Nina K. Jain, Pierce J. Ogden, Patrick F. Riley, George M. Church, Lucy J. Colwell, and Eric D. Kelsic. 2021. 'Deep Diversification of an AAV Capsid Protein by Machine Learning'. *Nature Biotechnology* 39 (6): 691–96. https://doi.org/10.1038/s41587-020-00793-4.

Bryksin, Anton V., and Ichiro Matsumura. 2010. 'Overlap Extension PCR Cloning: A Simple and Reliable Way to Create Recombinant Plasmids'. *BioTechniques* 48 (6): 463–65. https://doi.org/10.2144/000113418.

Buck, Stuart. 2021. 'Beware Performative Reproducibility'. *Nature* 595 (7866): 151–151. https://doi.org/10.1038/d41586-021-01824-z.

Budeski, Katherine. 2018. 'Biosecurity by Design: Getting Ahead of Risk in the World of Designer Organisms'. *The Nuclear Threat Initiative* (blog). 27 June 2018. https://www.nti.org/atomic-pulse/biosecurity-design-getting-ahead-risk-world-designer-organisms/.

Burki, Talha. 2018. 'Ban on Gain-of-Function Studies Ends'. *The Lancet Infectious Diseases* 18 (2): 148–49. https://doi.org/10.1016/S1473-3099(18)30006-9.

Burns, Cara C., Ousmane M. Diop, Roland W. Sutter, and Olen M. Kew. 2014. 'Vaccine-Derived Polioviruses'. *The Journal of Infectious Diseases* 210 (suppl_1): S283–93. https://doi.org/10.1093/infdis/jiu295.

Cai, Yizhi, Neta Agmon, Woo Jin Choi, Alba Ubide, Giovanni Stracquadanio, Katrina Caravelli, Haiping Hao, Joel S. Bader, and Jef D. Boeke. 2015. 'Intrinsic Biocontainment: Multiplex Genome Safeguards Combine Transcriptional and Recombinational Control of Essential Yeast Genes'. *Proceedings of the National Academy of Sciences* 112 (6): 1803–8. https://doi.org/10.1073/pnas.1424704112.

Calcedo, Roberto, Luk H. Vandenberghe, Guangping Gao, Jianping Lin, and James M. Wilson. 2009. 'Worldwide Epidemiology of Neutralizing Antibodies to Adeno-Associated Viruses'. *The Journal of Infectious Diseases* 199 (3): 381–90. https://doi.org/10.1086/595830.

California State Government. 2022. *AB-1963 California State University and University of California: Gene Synthesis Providers.* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202120220AB1963.

Callihan, Donald R., Marian Downing, Esmeralda Meyer, Luis Alberto Ochoa, Brian Petuch, Paul Tranchell, and David White. 2021. 'Considerations for Laboratory Biosafety and Biosecurity During the Coronavirus Disease 2019 Pandemic: Applying the ISO 35001:2019 Standard and High-Reliability Organizations Principles'. *Applied Biosafety* 26 (3): 113–22. https://doi.org/10.1089/apb.20.0068.

Carlson, Colin J. 2020. 'From PREDICT to Prevention, One Pandemic Later'. *The Lancet Microbe* 1 (1): e6–7. https://doi.org/10.1016/S2666-5247(20)30002-1.

Carlson, Colin J., Maxwell J. Farrell, Zoe Grange, Barbara A. Han, Nardus Mollentze, Alexandra L. Phelan, Angela L. Rasmussen, et al. 2021. 'The Future of Zoonotic Risk Prediction'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1837): 20200358. https://doi.org/10.1098/rstb.2020.0358.

Carlson, Colin J., Casey M. Zipfel, Romain Garnier, and Shweta Bansal. 2019. 'Global Estimates of Mammalian Viral Diversity Accounting for Host Sharing'. *Nature Ecology & Evolution* 3 (7): 1070–75. https://doi.org/10.1038/s41559-019-0910-6.

Carlson, Rob, and Rik Wehbring. 2020. 'Two Worlds, Two Bioeconomies: The Impacts of Decoupling US–China Trade and Technology Transfer'. Johns Hopkins Applied Physics Laboratory.

Carroll, Dennis, Peter Daszak, Nathan D. Wolfe, George F. Gao, Carlos M. Morel, Subhash Morzaria, Ariel Pablos-Méndez, Oyewale Tomori, and Jonna A. K. Mazet. 2018. 'The Global Virome Project'. *Science* 359 (6378): 872–74. https://doi.org/10.1126/science.aap7463.

Carus, Seth. 2017. *A Short History of Biological Warfare: From Pre-History to the 21 St Century*.

Casadevall, Arturo, Terence S. Dermody, Michael J. Imperiale, Rozanne M. Sandri-Goldin, and Thomas Shenk. 2014. 'On the Need for a National Board To Assess Dual Use Research of Concern'. *Journal of Virology* 88 (12): 6535–37. https://doi.org/10.1128/JVI.00875-14.

Casadevall, Arturo, and Michael J. Imperiale. 2014. 'Risks and Benefits of Gain-of-Function Experiments with Pathogens of Pandemic Potential, Such as Influenza Virus: A Call for a Science-Based Discussion'. *MBio* 5 (4). https://doi.org/10.1128/mBio.01730-14.

Casadevall, Arturo, and David A. Relman. 2010. 'Microbial Threat Lists: Obstacles in the Quest for Biosecurity?' *Nature Reviews. Microbiology* 8 (2): 149–54. https://doi.org/10.1038/nrmicro2299.

Casadevall, Arturo, and Thomas Shenk. 2012. 'The H5N1 Moratorium Controversy and Debate'. *MBio* 3 (5): e00379-12. https://doi.org/10.1128/mBio.00379-12.

Casey, Geoffrey A., Kimberly M. Papp, and Ian M. MacDonald. 2020. 'Ocular Gene Therapy with Adeno-Associated Virus Vectors: Current Outlook for Patients and Researchers'. *Journal of Ophthalmic and Vision Research (JOVR)*, July, 396–99. https://doi.org/10.18502/jovr.v15i3.7457.

Castle, Michael J., Heikki T. Turunen, Luk H. Vandenberghe, and John H. Wolfe. 2016. 'Controlling AAV Tropism in the Nervous System with Natural and Engineered Capsids'. In *Gene Therapy for Neurological Disorders: Methods and Protocols*, edited by Fredric P. Manfredsson, 133–49. Methods in Molecular Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3271-9_10.

Cello, Jeronimo, Aniko V. Paul, and Eckard Wimmer. 2002. 'Chemical Synthesis of Poliovirus CDNA: Generation of Infectious Virus in the Absence of Natural Template'. *Science* 297 (5583): 1016–18. https://doi.org/10.1126/science.1072266.

Centers for Disease Control and Prevention. 2014. 'Media Statement on Newly Discovered Smallpox Specimens'. CDC. 8 July 2014. https://www.cdc.gov/media/releases/2014/s0708-NIH.html.

———. 2019. 'Reconstruction of the 1918 Influenza Pandemic Virus'. 18 December 2019. https://www.cdc.gov/flu/about/qa/1918flupandemic.htm.

Centers for Disease Control and Prevention and U. S. Department of Agriculture. 2022. 'Federal Select Agent Program'. 2022. https://www.selectagents.gov/index.htm.

Central Intelligence Agency. 2005. *Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD, with Addendums (Duefler Report)*. Vol. 3. Washington, DC: U.S. Government Office. https://www.govinfo.gov/app/details/GPO-DUELFERREPORT.

Chaffin, Brian, Hannah Gosnell, and Barbara Cosens. 2014. 'A Decade of Adaptive Governance Scholarship: Synthesis and Future Directions'. *Ecology and Society* 19 (3). https://doi.org/10.5751/ES-06824-190356.

Chan, Ying Kai, Sean K. Wang, Colin J. Chu, David A. Copland, Alexander J. Letizia, Helena Costa Verdera, Jessica J. Chiang, et al. 2021. 'Engineering Adeno-Associated Viral Vectors to Evade Innate Immune and Inflammatory Responses'. *Science Translational Medicine* 13 (580). https://doi.org/10.1126/scitranslmed.abd3438.

Chen, Da-Yuan, Chue Vin Chin, Devin Kenney, Alexander H. Tavares, Nazimuddin Khan, Hasahn L. Conway, GuanQun Liu, et al. 2023. 'Spike and Nsp6 Are Key Determinants of SARS-CoV-2 Omicron BA.1 Attenuation'. *Nature*, January, 1–3. https://doi.org/10.1038/s41586-023-05697-2.

Chen, Da-Yuan, Devin Kenney, Chue Vin Chin, Alexander H. Tavares, Nazimuddin Khan, Hasahn L. Conway, GuanQun Liu, et al. 2022. 'Role of Spike in the Pathogenic and Antigenic Behavior of SARS-CoV-2 BA.1 Omicron'. bioRxiv. https://doi.org/10.1101/2022.10.13.512134.

Ciuffreda, Laura, Héctor Rodríguez-Pérez, and Carlos Flores. 2021. 'Nanopore Sequencing and Its Application to the Study of Microbial Communities'. *Computational and Structural Biotechnology Journal* 19 (January): 1497–1511. https://doi.org/10.1016/j.csbj.2021.02.020.

Clarke, Steve. 2013. 'The Precautionary Principle and the Dual-Use Dilemma'. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 223–34. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.19.

Coelho, Teresa, David Adams, Ana Silva, Pierre Lozeron, Philip N. Hawkins, Timothy Mant, Javier Perez, et al. 2013. 'Safety and Efficacy of RNAi Therapy for Transthyretin Amyloidosis'. *New England Journal of Medicine* 369 (9): 819–29. https://doi.org/10.1056/NEJMoa1208760.

Collins, Francis. 2021. 'Letter by National Institutes of Health to Grassley on COVID-19 Origins and Grant Oversight', 28 July 2021. https://www.grassley.senate.gov/imo/media/doc/national_institutes_of_health_to_grassley_-_covid_origins_grant_oversight.pdf.

Collins, H. 2010. *Tacit and Explicit Knowledge*. Chicago: ILUniversity of Chicago Press.

Corbett, Kizzmekia S., Darin K. Edwards, Sarah R. Leist, Olubukola M. Abiona, Seyhan Boyoglu-Barnum, Rebecca A. Gillespie, Sunny Himansu, et al. 2020. 'SARS-CoV-2 MRNA Vaccine Design Enabled by Prototype Pathogen Preparedness'. *Nature* 586 (August): 567-571(2020). https://doi.org/10.1038/s41586-020-2622-0.

Costa, Kevin, and Embriette Hyde. 2022. 'Cell-Free Expression Platforms Enable New Possibilities at IGEM and Beyond'. *Synbiobeta* (blog). 2022. https://www.synbiobeta.com/read/cell-free-expression-platforms-enable-new-possibilities-at-igem-and-beyond.

Cruickshank, Marilyn, and Ramon Z. Shaban. 2020. 'COVID-19: Lessons to Be Learnt from a Once-in-a-Century Global Pandemic'. *Journal of Clinical Nursing* 29 (21–22): 3901–4. https://doi.org/10.1111/jocn.15365.

Cullis, Pieter R., and Michael J. Hope. 2017. 'Lipid Nanoparticle Systems for Enabling Gene Therapies'. *Molecular Therapy* 25 (7): 1467–75. https://doi.org/10.1016/j.ymthe.2017.03.013.

Curry, Stephen, Sarah de Rijcke, Anna Hatch, Dorsamy (Gansen) Pillay, Inge van der Weijden, and James Wilsdon. 2020. 'The Changing Role of Funders in Responsible Research Assessment: Progress, Obstacles and the Way Ahead'. Report. Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v1.

Dafoe, Allan. 2015. 'On Technological Determinism: A Typology, Scope Conditions, and a Mechanism'. *Science, Technology, & Human Values* 40 (6): 1047–76.

Dalla-Torre, Hugo, Liam Gonzalez, Javier Mendoza-Revilla, Nicolas Lopez Carranza, Adam Henryk Grzywaczewski, Francesco Oteri, Christian Dallago, et al. 2023. 'The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics'. bioRxiv. https://doi.org/10.1101/2023.01.11.523679.

Dama, Adam C., Kevin S. Kim, Danielle M. Leyva, Annamarie P. Lunkes, Noah S. Schmid, Kenan Jijakli, and Paul A. Jensen. 2023. 'BacterAI Maps Microbial Metabolism without Prior Knowledge'. *Nature Microbiology*, May, 1–8. https://doi.org/10.1038/s41564-023-01376-0.

Danzig, Richard, Marc Sageman, Terrance Leighton, Lloyd Hough, Hidemi Yuki, Rui Kotani, and Zachary M Hosford. 2012. 'Aum Shinrikyo: Insights Into How Terrorists Develop Biological and Chemical Weapons'. Center for a New American Security.

Dauparas, J., I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, et al. 2022. 'Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN'. *Science* 378 (6615): 49–56. https://doi.org/10.1126/science.add2187.

Dean, Frank B., John R. Nelson, Theresa L. Giesler, and Roger S. Lasken. 2001. 'Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification'. *Genome Research* 11 (6): 1095–99. https://doi.org/10.1101/gr.180501.

Dechezleprêtre, Antoine, and Misato Sato. 2017. 'The Impacts of Environmental Regulations on Competitiveness'. *Review of Environmental Economics and Policy* 11 (2): 183–206. https://doi.org/10.1093/reep/rex013.

DeFrancesco, Laura. 2021. 'Synthetic Virology: The Experts Speak'. *Nature Biotechnology* 39 (10): 1185–93. https://doi.org/10.1038/s41587-021-01078-0.

Delebecque, Camille, and Jim Philip. 2015. 'Training for Synthetic Biology Jobs in the New Bioeconomy'. *Science Careers* (blog). 2015. https://www.science.org/content/article/training-

synthetic-biology-jobs-new-bioeconomy.

Denniston, Elizabeth, Hannah Crewdson, Nicole Rucinsky, Andrew Stegman, Diana Remenar, Katherine Moio, Brianne Clark, et al. 2016. 'The Practical Consideration of Poliovirus as an Oncolytic Virotherapy'. *American Journal of Virology* 5 (1): 1–7.

Department of State. 2016. '2016 Compliance Report'. https://2009-2017.state.gov/t/avc/rls/rpt/2016/index.htm.

Dhar, Pawan K, and Ron Weiss. 2007. 'Enabling the New Biology of the 21st Century'. *Systems and Synthetic Biology* 1 (1): 1–2. https://doi.org/10.1007/s11693-006-9000-6.

Dias, Marina França, Kwangsic Joo, Jessica A. Kemp, Silvia Ligório Fialho, Armando da Silva Cunha, Se Joon Woo, and Young Jik Kwon. 2018. 'Molecular Genetics and Emerging Therapies for Retinitis Pigmentosa: Basic Research and Clinical Perspectives'. *Progress in Retinal and Eye Research* 63 (March): 107–31. https://doi.org/10.1016/j.preteyeres.2017.10.004.

DiEuliis, Diane, and Charles Lutes. 2016. 'Security Implications of Emerging Biotechnologies: Workshop Summary, Analysis, and Recommendations'. National Defense University. https://wmdcenter.ndu.edu/Portals/97/Emerging%20Biotechnology%20Workshop%20Paper%20-%20April%202016.pdf.

Diggans, James, and Emily Leproust. 2019. 'Next Steps for Access to Safe, Secure DNA Synthesis'. *Frontiers in Bioengineering and Biotechnology* 7 (April): 86. https://doi.org/10.3389/fbioe.2019.00086.

Ding, Xiaozhe, and Viviana Gradinaru. 2020. 'Structure-Guided Rational Design of Adeno-Associated Viral Capsids with Expanded Sizes'. *Molecular Therapy* 28 (4): 226–27.

Dixon, Thomas A., Paul S. Freemont, Richard A. Johnson, and Isak S. Pretorius. 2022. 'A Global Forum on Synthetic Biology: The Need for International Engagement'. *Nature Communications* 13 (1): 3516. https://doi.org/10.1038/s41467-022-31265-9.

DNA Script. 2021. 'DNA Script Announces the Commercial Launch of the SYNTAX System, the First Benchtop DNA Printer Powered by Enzymatic Synthesis, to Accelerate Molecular Biology and Genomics Workflows'. *DNA Script* (blog). 15 June 2021. https://www.dnascript.com/press-releases/dna-script-announces-the-commercial-launch-of-the-syntax-system-the-first-benchtop-dna-printer-powered-by-enzymatic-synthesis-to-accelerate-molecular-biology-and-genomics-workflows/.

Douglas Friedman. 2018. 'One Percent'. Nuclear Threat Initiative. https://media.nti.org/documents/NTI_Initiative_-_Paper_2_-_Instituting_Biosecurity_Investment.pdf.

Douglas, Thomas. 2013. 'An Expected Value Approach to the Dual-Use Problem'. In *On the Dual Uses*

of Science and Ethics: Principles, Practices, and Prospects. ANU Press. https://philpapers.org/archive/DOUAEV.pdf.

Douglas, Thomas, and Julian Savulescu. 2010. 'Synthetic Biology and the Ethics of Knowledge'. *Journal of Medical Ethics* 36 (11): 687–93. https://doi.org/10.1136/jme.2010.038232.

Drobysz, Sonia. 2020. 'Verification and Implementation of the Biological and Toxin Weapons Convention'. *The Nonproliferation Review* 27 (4–6): 487–97. https://doi.org/10.1080/10736700.2020.1823102.

Dryzek, John S., and Aviezer Tucker. 2008. 'Deliberative Innovation to Different Effect: Consensus Conferences in Denmark, France, and the United States'. *Public Administration Review* 68 (5): 864–76.

Duprex, W. Paul, Ron A. M. Fouchier, Michael J. Imperiale, Marc Lipsitch, and David A. Relman. 2015. 'Gain-of-Function Experiments: Time for a Real Debate'. *Nature Reviews Microbiology* 13 (1): 58–64. https://doi.org/10.1038/nrmicro3405.

EASAC. 2015. 'Gain of Function: Experimental Applications Relating to Potentially Pandemic Pathogens'. https://www.degruyter.com/document/doi/10.1515/jwiet-2017-0017/html.

Edwards, Brett. 2019. *Insecurity and Emerging Biotechnology: Governing Misuse Potential*. 1st ed. 2019 edition. Palgrave Pivot.

Eisenstein, Michael. 2023. 'AI-Enhanced Protein Design Makes Proteins That Have Never Existed'. *Nature Biotechnology* 41 (3): 303–5. https://doi.org/10.1038/s41587-023-01705-y.

Elnaggar, Ahmed, Hazem Essam, Wafaa Salah-Eldin, Walid Moustafa, Mohamed Elkerdawy, Charlotte Rochereau, and Burkhard Rost. 2023. 'Ankh: Optimized Protein Language Model Unlocks General-Purpose Modelling'. arXiv. https://doi.org/10.48550/arXiv.2301.06568.

Elowitz, Michael B., and Stanislas Leibler. 2000. 'A Synthetic Oscillatory Network of Transcriptional Regulators'. *Nature* 403 (6767): 335–38. https://doi.org/10.1038/35002125.

Emerson, Claudia, Stephanie James, Katherine Littler, and Filippo (Fil) Randazzo. 2017. 'Principles for Gene Drive Research'. *Science* 358 (6367): 1135–36. https://doi.org/10.1126/science.aap9026.

Endy, Drew. 2005. 'Foundations for Engineering Biology'. *Nature* 438 (7067): 449–53. https://doi.org/10.1038/nature04342.

Enserink, Martin. 2004. 'Tiptoeing around Pandora's Box: Researchers Say Crossing Avian and Human Flu Viruses Is Crucial to Understanding the Threat of a New Influenza Pandemic, but They Admit That They Might Create a Monster'. *Science* 305 (5684): 594–96.

Esvelt, Kevin M. 2016. 'Gene Editing Can Drive Science to Openness'. *Nature* 534 (7606): 153–153.

https://doi.org/10.1038/534153a.

———. 2017. ‘Precaution: Open Gene Drive Research’. *Science*, February. https://www.science.org/doi/abs/10.1126/science.aal5325.

———. 2018. ‘Inoculating Science against Potential Pandemics and Information Hazards’. *PLoS Pathogens* 14 (10). https://doi.org/10.1371/journal.ppat.1007286.

Esvelt, Kevin M. 2022. ‘Delay, Detect, Defend: Preparing for a Future in Which Thousands Can Release New Pandemics’. Geneva Paper 29/22. Geneva Centre for Security Policy.

Esvelt, Kevin M., Andrea L Smidler, Flaminia Catteruccia, and George M Church. 2014. ‘Concerning RNA-Guided Gene Drives for the Alteration of Wild Populations’. Edited by Diethard Tautz. *ELife* 3 (July): e03401. https://doi.org/10.7554/eLife.03401.

European Commission. 2022. ‘Ethics Appraisal Procedure. Horizon 2020 Online Manual’. 2022. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm.

Evans, Nicholas G. 2013. ‘Contrasting Dual-Use Issues in Biology and Nuclear Science’. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 255–74. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.21.

———. 2014. ‘Dual-Use Decision Making: Relational and Positional Issues’. *Monash Bioethics Review* 32 (3): 268–83. https://doi.org/10.1007/s40592-015-0026-y.

———. 2018. ‘Ethical and Philosophical Considerations for Gain-of-Function Policy: The Importance of Alternate Experiments’. *Frontiers in Bioengineering and Biotechnology* 6. https://www.frontiersin.org/articles/10.3389/fbioe.2018.00011.

Evans, Sam Weiss. 2022. ‘When All Research Is Dual Use’. *Issues in Science and Technology* (blog). 23 May 2022. https://issues.org/dual-use-research-biosecurity-social-context-science-evans/.

Evans, Sam Weiss, Jacob Beal, Kavita Berger, Diederik A. Bleijs, Alessia Cagnetti, Francesca Ceroni, Gerald L. Epstein, et al. 2020. ‘Embrace Experimentation in Biosecurity Governance’. *Science* 368 (6487): 138–40. https://doi.org/10.1126/science.aba2932.

Evans, Sam Weiss, Daniel Greene, Connor Hoffmann, and Stefan Lunte. 2021. ‘Stakeholder Engagement Workshop on the Implementation of the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern: Workshop Report’. SSRN Scholarly Paper ID 3955051. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3955051.

Evans, Sam Weiss, and Megan J. Palmer. 2018. ‘Anomaly Handling and the Politics of Gene Drives’. *Journal of Responsible Innovation* 5 (sup1): S223–42.

https://doi.org/10.1080/23299460.2017.1407911.

Falkow, Stanley. 2012. 'The Lessons of Asilomar and the H5N1 "Affair"'. *MBio* 3 (5): e00354-12. https://doi.org/10.1128/mBio.00354-12.

Falsey, Ann R., Magdalena E. Sobieszczyk, Ian Hirsch, Stephanie Sproule, Merlin L. Robb, Lawrence Corey, Kathleen M. Neuzil, et al. 2021. 'Phase 3 Safety and Efficacy of AZD1222 (ChAdOx1 NCoV-19) Covid-19 Vaccine'. *New England Journal of Medicine* 385 (25): 2348–60. https://doi.org/10.1056/NEJMoa2105290.

Fan, Yongfeng, Jason R. Barash, Jianlong Lou, Fraser Conrad, James D. Marks, and Stephen S. Arnon. 2016. 'Immunological Characterization and Neutralizing Ability of Monoclonal Antibodies Directed Against Botulinum Neurotoxin Type H'. *The Journal of Infectious Diseases* 213 (10): 1606–14. https://doi.org/10.1093/infdis/jiv770.

Farquhar, Sebastian, Owen Cotton-Barratt, and Andrew Snyder-Beattie. 2017. 'Pricing Externalities to Balance Public Risks and Benefits of Research'. *Health Security* 15 (4): 401–8. https://doi.org/10.1089/hs.2016.0118.

Fears, Robin, and Volker ter Meulen. 2015. 'What next for Gain-of-Function Research in Europe?' *ELife* 4: e13035. https://doi.org/10.7554/eLife.13035.

Federal Office of Consumer Protection and Food Safety. 2022. 'Genetic Engineering Act'. 2022. https://www.bvl.bund.de/EN/Tasks/06_Genetic_engineering/02_Consumers/07_Legal_Framework/01_Germany/Germany_node.html.

Federal Register. 2020. 'Review and Revision of the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA'. *Federal Register*, 26 August 2020. https://www.federalregister.gov/documents/2020/08/26/2020-18444/review-and-revision-of-the-screening-framework-guidance-for-providers-of-synthetic-double-stranded.

Felbinger, Jonathan, and Judith Reppy. 2011. 'Classifying Knowledge, Creating Secrets: Government Policy for Dual-Use Technology'. In *Government Secrecy*, edited by Susan Maret, 19:277–99. Research in Social Problems and Public Policy. Emerald Group Publishing Limited. https://doi.org/10.1108/S0196-1152(2011)0000019019.

Finkel, Elizabeth. 2001. 'Engineered Mouse Virus Spurs Bioweapon Fears'. *Science* 291 (5504): 585–585. https://doi.org/10.1126/science.291.5504.585.

Florea, Michael, Henrik Hagemann, Gabriella Santosa, James Abbott, Chris N. Micklem, Xenia Spencer-Milnes, Laura de Arroyo Garcia, et al. 2016. 'Engineering Control of Bacterial Cellulose Production Using a Genetic Toolkit and a New Cellulose-Producing Strain'. *Proceedings of the National Academy of Sciences* 113 (24): E3431–40. https://doi.org/10.1073/pnas.1522985113.

Folegatti, Pedro M, Katie J Ewer, Parvinder K Aley, Brian Angus, Stephan Becker, Sandra Belij-Rammerstorfer, Duncan Bellamy, et al. 2020. 'Safety and Immunogenicity of the ChAdOx1 NCoV-19 Vaccine against SARS-CoV-2: A Preliminary Report of a Phase 1/2, Single-Blind, Randomised Controlled Trial'. *The Lancet* 0 (0). https://doi.org/10.1016/S0140-6736(20)31604-4.

Forge, John. 2013. 'Responsible Dual Use'. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 121–32. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.13.

Fouchier, Ron A. M., Adolfo García-Sastre, Yoshihiro Kawaoka, Wendy S. Barclay, Nicole M. Bouvier, Ian H. Brown, Ilaria Capua, et al. 2012. 'Pause on Avian Flu Transmission Research'. *Science* 335 (6067): 400–401. https://doi.org/10.1126/science.1219412.

Franco, Crystal, and Tara Kirk Sell. 2011. 'Federal Agency Biodefense Funding, FY2011-FY2012'. *Biosecurity and Bioterrorism* 9 (2): 22. https://doi.org/10.1089/bsp.2011.0018.

Fredens, Julius, Kaihang Wang, Daniel de la Torre, Louise F. H. Funke, Wesley E. Robertson, Yonka Christova, Tiongsun Chia, et al. 2019. 'Total Synthesis of Escherichia Coli with a Recoded Genome'. *Nature* 569 (7757): 514–18. https://doi.org/10.1038/s41586-019-1192-5.

Fuhrmann, Matthew. 2012. 'Spreading Temptation: Why Nuclear Export Strategies Backfire'. In *Atomic Assistance: How 'Atoms for Peace' Programs Cause Nuclear Insecurity*, edited by Kevin P. Gallagher, 0. Cornell University Press. https://doi.org/10.7591/cornell/9780801450907.003.0008.

Galdzicki, Michal, Cesar Rodriguez, Deepak Chandran, Herbert M. Sauro, and John H. Gennari. 2011. 'Standard Biological Parts Knowledgebase'. *PLoS ONE* 6 (2): e17005. https://doi.org/10.1371/journal.pone.0017005.

Galič, Maša, Tjerk Timan, and Bert-Jaap Koops. 2017. 'Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation'. *Philosophy & Technology* 30 (1): 9–37. https://doi.org/10.1007/s13347-016-0219-1.

Gardner, Timothy S., Charles R. Cantor, and James J. Collins. 2000. 'Construction of a Genetic Toggle Switch in Escherichia Coli'. *Nature* 403 (6767): 339–42. https://doi.org/10.1038/35002131.

Garrie, Daniel, and Michael Mann. 2014. 'Cyber-Security Insurance: Navigating the Landscape of a Growing Field'. *UIC John Marshall Journal of Information Technology & Privacy Law* 31 (3). https://repository.law.uic.edu/jitpl/vol31/iss3/4.

Ge, Xing-Yi, Jia-Lu Li, Xing-Lou Yang, Aleksei A. Chmura, Guangjian Zhu, Jonathan H. Epstein, Jonna K. Mazet, et al. 2013. 'Isolation and Characterization of a Bat SARS-like Coronavirus That Uses the ACE2 Receptor'. *Nature* 503 (7477): 535–38. https://doi.org/10.1038/nature12711.

Geels, Frank W., Benjamin K. Sovacool, Tim Schwanen, and Steve Sorrell. 2017. 'Sociotechnical Transitions for Deep Decarbonization'. *Science* 357 (6357): 1242–44. https://doi.org/10.1126/science.aao3760.

Geoghegan, Jemma L., and Edward C. Holmes. 2017. 'Predicting Virus Emergence amid Evolutionary Noise'. *Open Biology* 7 (10): 170189. https://doi.org/10.1098/rsob.170189.

German Ethics Council. 2014. 'Biosecurity — Freedom and Responsibility of Research'. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/englisch/opinion-biosecurity.pdf.

German Federal Foreign Office. 2022. 'BWC Confidence Building Measure'. 2022. https://bwc-ecbm.unog.ch/germany/bwccbm2022germany.

Gibson, Daniel G., Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, et al. 2008. 'Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma Genitalium Genome'. *Science (New York, N.Y.)* 319 (5867): 1215–20. https://doi.org/10.1126/science.1151721.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. 2009. 'Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases'. *Nature Methods* 6 (5): 343–45. https://doi.org/10.1038/nmeth.1318.

Gilsdorf, Janet R., and Raymond A. Zilinskas. 2005. 'New Considerations in Infectious Disease Outbreaks: The Threat of Genetically Modified Microbes'. *Clinical Infectious Diseases* 40 (8): 1160–65. https://doi.org/10.1086/428843.

Global Partnership Against the Spread of Weapons and Material of Mass Destruction. 2022. 'Africa Signature Initiative'. 2022. https://www.gpwmd.com/africa-signature-intitiative.

Goldstein, Josh A., Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, and Katerina Sedova. 2023. 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations'. arXiv. http://arxiv.org/abs/2301.04246.

Gould, Chandré, Peter Folb, Peter I. Folb, United Nations, United Nations Institute for Disarmament Research Staff, and Centre for Conflict Resolution-Kenya Staff. 2002. *Project Coast: Apartheid's Chemical and Biological Warfare Programme*. United Nations Institute for Disarmament Research (UNIDIR).

Government of Canada. 2014. 'Human Pathogens and Toxins Regulations'. *Canada Gazette*, 21 June 2014. https://gazette.gc.ca/rp-pr/p1/2014/2014-06-21/html/reg2-eng.html.

Government of the United Kingdom. 2001. 'Anti-Terrorism, Crime and Security Act 2001'. Text. Statute Law Database. https://www.legislation.gov.uk/ukpga/2001/24/schedule/5.

Graham, Barney S., and Nancy J. Sullivan. 2018. 'Emerging Viral Diseases from a Vaccinology Perspective: Preparing for the next Pandemic'. *Nature Immunology* 19 (1): 20–28. https://doi.org/10.1038/s41590-017-0007-9.

Grange, Zoë L., Tracey Goldstein, Christine K. Johnson, Simon Anthony, Kirsten Gilardi, Peter Daszak, Kevin J. Olival, et al. 2021. 'Ranking the Risk of Animal-to-Human Spillover for Newly Discovered Viruses'. *Proceedings of the National Academy of Sciences* 118 (15). https://doi.org/10.1073/pnas.2002324118.

Gray, Gregory C., Emily R. Robie, Caleb J. Studstill, and Charles L. Nunn. 2021. 'Mitigating Future Respiratory Virus Pandemics: New Threats and Approaches to Consider'. *Viruses* 13 (4): 637. https://doi.org/10.3390/v13040637.

Green, Jessica F. 2021. 'Does Carbon Pricing Reduce Emissions? A Review of Ex-Post Analyses'. *Environmental Research Letters* 16 (4): 043004. https://doi.org/10.1088/1748-9326/abdae9.

Gretton, Dana, Erika A. DeBenedicts, Andrew B. Liu, Andrew C. Yao, and Kevin M. Esvelt. 2021. 'Random Adversarial Threshold Search Enables Specific, Secure, and Automated DNA Synthesis Screening'. https://www.securedna.org/download/Random_Adversarial_Threshold_Screening.pdf.

Griffiths, Megan E., Alice Broos, Laura M. Bergner, Diana K. Meza, Nicolas M. Suarez, Ana da Silva Filipe, Carlos Tello, Daniel J. Becker, and Daniel G. Streicker. 2022. 'Longitudinal Deep Sequencing Informs Vector Selection and Future Deployment Strategies for Transmissible Vaccines'. *PLOS Biology* 20 (4): e3001580. https://doi.org/10.1371/journal.pbio.3001580.

Gronvall, Gigi. 2016. *Synthetic Biology: Safety, Security, and Promise*. Health Security Press.

Gronvall, Gigi Kwik. 2018. 'Safety, Security, and Serving the Public Interest in Synthetic Biology'. *Journal of Industrial Microbiology and Biotechnology* 45 (7): 463–66. https://doi.org/10.1007/s10295-018-2026-4.

Gryphon Scientific. 2016. 'Risk and Benefit Analysis of Gain of Function Research'.

Guillemin, Jeanne. 2011. *American Anthrax: Fear, Crime, and the Investigation of the Nation's Deadliest Bioterror Attack*. New York: Times Books.

Gupta, Joyeeta. 2010. 'A History of International Climate Change Policy'. *WIREs Climate Change* 1 (5): 636–53. https://doi.org/10.1002/wcc.67.

Haberman, R. P., T. J. McCown, and R. J. Samulski. 1998. 'Inducible Long-Term Gene Expression in Brain with Adeno-Associated Virus Gene Transfer'. *Gene Therapy* 5 (12): 1604–11. https://doi.org/10.1038/sj.gt.3300782.

Haegel, Nancy M., Harry Atwater, Teresa Barnes, Christian Breyer, Anthony Burrell, Yet-Ming Chiang,

Stefaan De Wolf, et al. 2019. ‘Terawatt-Scale Photovoltaics: Transform Global Energy’. *Science* 364 (6443): 836–38. https://doi.org/10.1126/science.aaw1845.

Haegel, Nancy M., Robert Margolis, Tonio Buonassisi, David Feldman, Armin Froitzheim, Raffi Garabedian, Martin Green, et al. 2017. ‘Terawatt-Scale Photovoltaics: Trajectories and Challenges’. *Science* 356 (6334): 141–43. https://doi.org/10.1126/science.aal1288.

Hanson, David. 1992. ‘Toxics Release Inventory Data Show Steady Drop in Emissions’. *Chemical & Engineering News*. https://doi.org/10.1021/cen-v070n024.p013.

Harari, Daniel, Matthew Keep, and Philip Brien. 2022. ‘Coronavirus: Economic Impact’, December. https://commonslibrary.parliament.uk/research-briefings/cbp-8866/.

Hardin, Garrett. 1968. ‘The Tragedy of the Commons’. *Science* 162 (3859): 1243–48.

Harris, Elisa. 2020. ‘North Korea and Biological Weapons: Assessing the Evidence’. *Stimson Center* (blog). 6 November 2020. https://www.stimson.org/2020/north-korea-and-biological-weapons-assessing-the-evidence/.

Hayden, Erika Check. 2011. ‘Bioengineers Debate Use of Military Money’. *Nature* 479 (7374): 458–458. https://doi.org/10.1038/479458a.

Health and Safety Executive. 2021. ‘The SACGM Compendium of Guidance’. 2021. https://www.hse.gov.uk/biosafety/gmo/acgm/acgmcomp/.

Henao-Restrepo, Ana Maria, Anton Camacho, Ira M. Longini, Conall H. Watson, W. John Edmunds, Matthias Egger, Miles W. Carroll, et al. 2017. ‘Efficacy and Effectiveness of an RVSV-Vectored Vaccine in Preventing Ebola Virus Disease: Final Results from the Guinea Ring Vaccination, Open-Label, Cluster-Randomised Trial (Ebola Ça Suffit!)’. *The Lancet* 389 (10068): 505–18. https://doi.org/10.1016/S0140-6736(16)32621-6.

Henkel, Richard D., Thomas Miller, and Robbin S. Weyant. 2012. ‘Monitoring Select Agent Theft, Loss and Release Reports in the United States—2004–2010’. *Applied Biosafety* 17 (4): 171–80. https://doi.org/10.1177/153567601201700402.

Herfst, Sander, Eefje J. A. Schrauwen, Martin Linster, Salin Chutinimitkul, Emmie de Wit, Vincent J. Munster, Erin M. Sorrell, et al. 2012. ‘Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets’. *Science* 336 (6088): 1534–41. https://doi.org/10.1126/science.1213362.

Hodgson, Andrea, Mary Maxon, and Joe Alper. 2022. ‘The U.S. Bioeconomy: Charting a Course for a Resilient and Competitive Future’. New York, NY: Schmidt Futures. https://doi.org/10.55879/d2hrs7zwc.

Holmes, Edward C., Andrew Rambaut, and Kristian G. Andersen. 2018. ‘Pandemics: Spend on Surveillance, Not Prediction’. *Nature* 558 (7709): 180–82. https://doi.org/10.1038/d41586-

018-05373-w.

Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. 'LoRA: Low-Rank Adaptation of Large Language Models'. arXiv. https://doi.org/10.48550/arXiv.2106.09685.

Huang, Sally. 2021. 'Commentary – Assessing China's New Biosafety Law'. *The Pandora Report* (blog). 19 March 2021. https://pandorareport.org/2021/03/19/commentary-assessing-chinas-new-biosafety-law/.

Hubrecht, Robert C., and Elizabeth Carter. 2019. 'The 3Rs and Humane Experimental Technique: Implementing Change'. *Animals : An Open Access Journal from MDPI* 9 (10): 754. https://doi.org/10.3390/ani9100754.

Hughes, Randall A., and Andrew D. Ellington. 2017. 'Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology'. *Cold Spring Harbor Perspectives in Biology* 9 (1): a023812. https://doi.org/10.1101/cshperspect.a023812.

Imai, Masaki, Tokiko Watanabe, Masato Hatta, Subash C. Das, Makoto Ozawa, Kyoko Shinya, Gongxun Zhong, et al. 2012. 'Experimental Adaptation of an Influenza H5 HA Confers Respiratory Droplet Transmission to a Reassortant H5 HA/H1N1 Virus in Ferrets'. *Nature* 486 (7403): 420–28. https://doi.org/10.1038/nature10831.

Imperiale, Michael J., and Arturo Casadevall. 2018. 'A New Approach to Evaluating the Risk–Benefit Equation for Dual-Use and Gain-of-Function Research of Concern'. *Frontiers in Bioengineering and Biotechnology* 6. https://doi.org/10.3389/fbioe.2018.00021.

Inagaki, Takashi, Akari Kato, Koichi Takahashi, Haruka Ozaki, and Genki N. Kanda. 2023. 'LLMs Can Generate Robotic Scripts from Goal-Oriented Instructions in Biological Laboratory Automation'. arXiv. https://doi.org/10.48550/arXiv.2304.10267.

Inglesby, Thomas V. 2018. 'Horsepox and the Need for a New Norm, More Transparency, and Stronger Oversight for Experiments That Pose Pandemic Risks'. *PLoS Pathogens* 14 (10). https://doi.org/10.1371/journal.ppat.1007129.

Inglesby, Thomas V., and Marc Lipsitch. 2020. 'Proposed Changes to U.S. Policy on Potential Pandemic Pathogen Oversight and Implementation'. *MSphere* 5 (1). https://doi.org/10.1128/mSphere.00990-19.

Inglesby, Thomas V., and David A Relman. 2016. 'How Likely Is It That Biological Agents Will Be Used Deliberately to Cause Widespread Harm?' *EMBO Reports* 17 (2): 127–30. https://doi.org/10.15252/embr.201541674.

International Gene Synthesis Consortium. 2017. 'Harmonized Screening Protocol v2.0'. International Gene Synthesis Corporation. https://genesynthesisconsortium.org/wp-

content/uploads/IGSCHarmonizedProtocol11-21-17.pdf.

Isaac, Christopher R. 2022. 'Establishing an Incentive-Based Multistakeholder Approach to Dual-Use DNA Screening'. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire* 100 (3): 268–73. https://doi.org/10.1139/bcb-2021-0504.

Jackson, Ronald J., Alistair J. Ramsay, Carina D. Christensen, Sandra Beaton, Diana F. Hall, and Ian A. Ramshaw. 2001. 'Expression of Mouse Interleukin-4 by a Recombinant Ectromelia Virus Suppresses Cytolytic Lymphocyte Responses and Overcomes Genetic Resistance to Mousepox'. *Journal of Virology* 75 (3): 1205–10. https://doi.org/10.1128/JVI.75.3.1205-1210.2001.

Jackson, Shawn S., Louise E. Sumner, Christian H. Garnier, Casey Basham, Landy T. Sun, Peter L. Simone, Danielle S. Gardner, and Rocco J. Casagrande. 2019. 'The Accelerating Pace of Biotech Democratization'. *Nature Biotechnology* 37 (12): 1403–8. https://doi.org/10.1038/s41587-019-0339-0.

Jacobsen, Kirsten X., Kirsten Mattison, Marianne Heisz, and Sandra Fry. 2014. 'Biosecurity in Emerging Life Sciences Technologies, a Canadian Public Health Perspective'. *Frontiers in Public Health* 2. https://doi.org/10.3389/fpubh.2014.00198.

Jasanoff, Sheila. 2007. 'Technologies of Humility'. *Nature* 450 (7166): 33–33. https://doi.org/10.1038/450033a.

Javitt, Gail, and Anya Prince. 2012. 'Gene Therapy'. In *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies*, 249–59. Cambridge, MA: MIT Press.

Jefferson, Catherine, Filippa Lentzos, and Claire Marris. 2014. 'Synthetic Biology and Biosecurity: Challenging the "Myths"'. *Frontiers in Public Health* 2. https://www.frontiersin.org/articles/10.3389/fpubh.2014.00115.

Jeune, Vedell Louis, Jakob A. Joergensen, Roger J. Hajjar, and Thomas Weber. 2013. 'Pre-Existing Anti–Adeno-Associated Virus Antibodies as a Challenge in AAV Gene Therapy'. *Human Gene Therapy Methods* 24 (2): 59–67. https://doi.org/10.1089/hgtb.2012.243.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science (New York, N.Y.)* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Journal Editors. 2003. 'Statement on Scientific Publication and Security'. *Science* 299 (5610): 1149–1149. https://doi.org/10.1126/science.299.5610.1149.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. 'Highly Accurate Protein Structure Prediction with

AlphaFold'. *Nature* 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Kaiser, Jocelyn. 2021. 'NIH Says Grantee Failed to Report Experiment in Wuhan That Created a Bat Virus That Made Mice Sicker'. *Science Insider*, 21 October 2021. https://www.science.org/content/article/nih-says-grantee-failed-report-experiment-wuhan-created-bat-virus-made-mice-sicker.

Kaplan, C. 1989. 'Vaccinia Virus: A Suitable Vehicle for Recombinant Vaccines?' *Archives of Virology* 106 (1): 127–39. https://doi.org/10.1007/BF01311044.

Kaufman, Noah. 2016. 'Carbon Tax vs. Cap-and-Trade: What's a Better Policy to Cut Emissions?', January. https://www.wri.org/insights/carbon-tax-vs-cap-and-trade-whats-better-policy-cut-emissions.

Keiichi, Tsuneishi, and John Junkerman. 2013. 'Unit 731 and the Japanese Imperial Army's Biological Warfare Program'. In *Japan's Wartime Medical Atrocities: Comparative Inquiries in Science, History, and Ethics*, 21–31. Taylor and Francis. https://doi.org/10.4324/9780203849040.

Kemp, Luke, Laura Adam, Christian R Boehm, Rainer Breitling, Rocco Casagrande, Malcolm Dando, Appolinaire Djikeng, et al. 2020. 'Bioengineering Horizon Scan 2020'. Edited by Helena Pérez Valle, Peter Rodgers, and Ariel B Lindner. *ELife* 9 (May): e54489. https://doi.org/10.7554/eLife.54489.

Kemp, Luke, Chi Xu, Joanna Depledge, Kristie L. Ebi, Goodwin Gibbins, Timothy A. Kohler, Johan Rockström, et al. 2022. 'Climate Endgame: Exploring Catastrophic Climate Change Scenarios'. *Proceedings of the National Academy of Sciences* 119 (34): e2108146119. https://doi.org/10.1073/pnas.2108146119.

Kennedy, Donald. 2005. 'Better Never Than Late'. *Science* 310 (5746): 195–195. https://doi.org/10.1126/science.310.5746.195.

Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, et al. 2016. 'Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency'. *PLOS Biology* 14 (5): e1002456. https://doi.org/10.1371/journal.pbio.1002456.

Kilobaser. 2022. 'DNA & RNA Synthesizer'. Kilobaser - Personal DNA/RNA Synthesizer. 2022. https://kilobaser.com/.

Kim, M., M. Jeong, S. Hur, Y. Cho, J. Park, H. Jung, Y. Seo, et al. 2021. 'Engineered Ionizable Lipid Nanoparticles for Targeted Delivery of RNA Therapeutics into Different Types of Cells in the Liver'. *Science Advances* 7 (9): eabf4398. https://doi.org/10.1126/sciadv.abf4398.

Kirkpatrick, Jesse, Gregory D. Koblentz, Megan J. Palmer, Edward Perello, David A. Relman, and Sarah W. Denton. 2018. 'Editing Biosecurity: Needs and Strategies for Governing Genome Editing'.

http://mars.gmu.edu/bitstream/handle/1920/11342/Editing-Bio-%2bReport-Final.pdf?sequence=1&isAllowed=y.

Kitcher, Philip. 2003. *Science, Truth, and Democracy*. OUP USA.

Knight, Thomas. 2003. 'Idempotent Vector Design for Standard Assembly of Biobricks'. https://dspace.mit.edu/handle/1721.1/21168.

Koblentz, Gregory D. 2017. 'The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox'. *Health Security* 15 (6): 620–28. https://doi.org/10.1089/hs.2017.0061.

———. 2018. 'A Critical Analysis of the Scientific and Commercial Rationales for the De Novo Synthesis of Horsepox Virus'. *MSphere* 3 (2). https://doi.org/10.1128/mSphere.00040-18.

Kosuri, Sriram, and George M Church. 2014. 'Large-Scale de Novo DNA Synthesis: Technologies and Applications'. *Nature Methods* 11 (5): 499–507. https://doi.org/10.1038/nmeth.2918.

Kou, Zheng, Yi-Fan Huang, Ao Shen, Saeed Kosari, Xiang-Rong Liu, and Xiao-Li Qiang. 2021. 'Prediction of Pandemic Risk for Animal-Origin Coronavirus Using a Deep Learning Method'. *Infectious Diseases of Poverty* 10 (1): 128. https://doi.org/10.1186/s40249-021-00912-6.

Kuhlau, Frida, Anna T. Höglund, Kathinka Evers, and Stefan Eriksson. 2011. 'A Precautionary Principle for Dual Use Research in the Life Sciences'. *Bioethics* 25 (1): 1–8. https://doi.org/10.1111/j.1467-8519.2009.01740.x.

Kuiken, Todd. 2016. 'Governance: Learn from DIY Biologists'. *Nature* 531 (7593): 167–68. https://doi.org/10.1038/531167a.

Kunz, Alexis H., and Dieter Pfaff. 2002. 'Agency Theory, Performance Evaluation, and the Hypothetical Construct of Intrinsic Motivation'. *Accounting, Organizations and Society* 27 (3): 275–95. https://doi.org/10.1016/S0361-3682(01)00031-9.

Kupferschmidt, Kai. 2017. 'How Canadian Researchers Reconstituted an Extinct Poxvirus for $100,000 Using Mail-Order DNA'. *Science Insider*, July. https://doi.org/doi: 10.1126/science.aan7069.

———. 2018. 'A Paper Showing How to Make a Smallpox Cousin Just Got Published. Critics Wonder Why'. *ScienceInsider*, January. https://doi.org/doi: 10.1126/science.aat0607.

Ladner, Jason T. 2021. 'Genomic Signatures for Predicting the Zoonotic Potential of Novel Viruses'. *PLOS Biology* 19 (9): e3001403. https://doi.org/10.1371/journal.pbio.3001403.

Lanouette, William. 1992. *Genius in the Shadows: A Biography of Leo Szilard : The Man behind the Bomb*. New York : Oxford: Charles Scribner's Sons ; Maxwell Macmillan International.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Edited by

Jonas Salk. 2nd ed. edition. Princeton, N.J: Princeton University Press.

Lee, Thomas H. 2015. 'Financial versus Non-Financial Incentives for Improving Patient Experience'. *Journal of Patient Experience* 2 (1): 4–6. https://doi.org/10.1177/237437431500200102.

Leitenberg, Milton. 1999. 'Aum Shinrikyo's Efforts to Produce Biological Weapons: A Case Study in the Serial Propagation of Misinformation'. *Terrorism and Political Violence* 11 (4): 149–58. https://doi.org/10.1080/09546559908427537.

———. 2001. 'Biological Weapons in the Twentieth Century: A Review and Analysis'. *Critical Reviews in Microbiology* 27 (4): 267–320. https://doi.org/10.1080/20014091096774.

———. 2005. 'Assessing the Biological Weapons and Bioterrorism Threat', January.

Leitenberg, Milton, and Raymond A. Zilinskas. 2012. *The Soviet Biological Weapons Program: A History*. Cambridge: Harvard University Press.

Leman, Julia Koehler, Brian D. Weitzner, Steven M. Lewis, Jared Adolf-Bryfogle, Nawsad Alam, Rebecca F. Alford, Melanie Aprahamian, et al. 2020. 'Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks'. *Nature Methods* 17 (7): 665–80. https://doi.org/10.1038/s41592-020-0848-2.

Lentzos, Filippa. 2019. 'Compliance and Enforcement in the Biological Weapons Regime'. WMDCE Series No. 4. Geneva, Switzerland: UNIDIR. https://doi.org/10.37559/WMD/19/WMDCE4.

Lentzos, Filippa, Edward P. Rybicki, Margret Engelhard, Pauline Paterson, Wayne Arthur Sandholtz, and R. Guy Reeves. 2022. 'Eroding Norms over Release of Self-Spreading Viruses'. *Science* 375 (6576): 31–33. https://doi.org/10.1126/science.abj5593.

Lev, Ori. 2019. 'Regulating Dual-Use Research: Lessons from Israel and the United States'. *Journal of Biosafety and Biosecurity* 1 (2): 80–85. https://doi.org/10.1016/j.jobb.2019.06.001.

Lewis, Gregory, Piers Millett, Anders Sandberg, Andrew Snyder-Beattie, and Gigi Gronvall. 2019. 'Information Hazards in Biotechnology'. *Risk Analysis* 39 (5): 975–81. https://doi.org/10.1111/risa.13235.

Li, F. C. K., B. C. K. Choi, T. Sly, and A. W. P. Pak. 2008. 'Finding the Real Case-Fatality Rate of H5N1 Avian Influenza'. *Journal of Epidemiology & Community Health* 62 (6): 555–59. https://doi.org/10.1136/jech.2007.064030.

Lin, Jianping, Roberto Calcedo, Luk H. Vandenberghe, Peter Bell, Suryanarayan Somanathan, and James M. Wilson. 2009. 'A New Genetic Vaccine Platform Based on an Adeno-Associated Virus Isolated from a Rhesus Macaque'. *Journal of Virology* 83 (24): 12738–50. https://doi.org/10.1128/JVI.01441-09.

Lipsitch, Marc. 2014. 'Can Limited Scientific Value of Potential Pandemic Pathogen Experiments Justify the Risks?' *MBio* 5 (5): e02008-14. https://doi.org/10.1128/mBio.02008-14.

Liu, Gary, Denise B. Catacutan, Khushi Rathod, Kyle Swanson, Wengong Jin, Jody C. Mohammed, Anush Chiappino-Pepe, et al. 2023. 'Deep Learning-Guided Discovery of an Antibiotic Targeting Acinetobacter Baumannii'. *Nature Chemical Biology*, May, 1–9. https://doi.org/10.1038/s41589-023-01349-8.

Lochrie, Michael A., Gwen P. Tatsuno, Brian Christie, Jennifer Wellman McDonnell, Shangzhen Zhou, Richard Surosky, Glenn F. Pierce, and Peter Colosi. 2006. 'Mutations on the External Surfaces of Adeno-Associated Virus Type 2 Capsids That Affect Transduction and Neutralization'. *Journal of Virology* 80 (2): 821–34. https://doi.org/10.1128/JVI.80.2.821-834.2006.

Lunshof, Jeantine E., and Angela Birnbaum. 2017. 'Adaptive Risk Management of Gene Drive Experiments: Biosafety, Biosecurity, and Ethics'. *Applied Biosafety* 22 (3): 97–103. https://doi.org/10.1177/1535676017721488.

Luo, Le, Yi-Chen Lan, and Qingliang Tang. 2012. 'Corporate Incentives to Disclose Carbon Information: Evidence from the CDP Global 500 Report'. *Journal of International Financial Management & Accounting* 23 (2): 93–120. https://doi.org/10.1111/j.1467-646X.2012.01055.x.

Luo, Renqian, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. 'BioGPT: Generative Pre-Trained Transformer for Biomedical Text Generation and Mining'. *Briefings in Bioinformatics* 23 (6): bbac409. https://doi.org/10.1093/bib/bbac409.

Lynch, Michael. 2002. 'Protocols, Practices, and the Reproduction of Technique in Molecular Biology*'. *The British Journal of Sociology* 53 (2): 203–20. https://doi.org/10.1080/00071310220133304.

Ma, Siying, Ishtiaq Saaem, and Jingdong Tian. 2012. 'Error Correction in Gene Synthesis Technology'. *Trends in Biotechnology* 30 (3): 147–54. https://doi.org/10.1016/j.tibtech.2011.10.002.

MacIntyre, Chandini Raina, Dillon Charles Adam, Robin Turner, Abrar Ahmad Chughtai, and Thomas Engells. 2020. 'Public Awareness, Acceptability and Risk Perception about Infectious Diseases Dual-Use Research of Concern: A Cross-Sectional Survey'. *BMJ Open* 10 (1): e029134. https://doi.org/10.1136/bmjopen-2019-029134.

MacKenzie, Donald, and Graham Spinardi. 1995. 'Tacit Knowledge, Weapons Design, and the Uninvention of Nuclear Weapons'. *American Journal of Sociology* 101 (1): 44–99. https://doi.org/10.1086/230699.

Madani, Ali, Ben Krause, Eric R. Greene, Subu Subramanian, Benjamin P. Mohr, James M. Holton, Jose Luis Olmos, et al. 2023. 'Large Language Models Generate Functional Protein Sequences across Diverse Families'. *Nature Biotechnology*, January, 1–8. https://doi.org/10.1038/s41587-022-

01618-2.

Maersch, Stephan, Anke Huber, Hildegard Büning, Michael Hallek, and Luca Perabo. 2010. 'Optimization of Stealth Adeno-Associated Virus Vectors by Randomization of Immunogenic Epitopes'. *Virology* 397 (1): 167–75. https://doi.org/10.1016/j.virol.2009.10.021.

Mahfoud, Tara, Christine Aicardi, Saheli Datta, and Nikolas Rose. 2018. 'The Limits of Dual Use'. *Issues in Science and Technology* (blog). 31 July 2018. https://issues.org/the-limits-of-dual-use/.

Making, Institute of Medicine (US) Committee to Study Decision, and Kathi E. Hanna. 1991. *Asilomar and Recombinant DNA: The End of the Beginning*. *Biomedical Politics*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK234217/.

Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. 'RNA-Guided Human Genome Engineering via Cas9'. *Science (New York, N.Y.)* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Mandell, Daniel J., Marc J. Lajoie, Michael T. Mee, Ryo Takeuchi, Gleb Kuznetsov, Julie E. Norville, Christopher J. Gregg, Barry L. Stoddard, and George M. Church. 2015. 'Biocontainment of Genetically Modified Organisms by Synthetic Protein Design'. *Nature* 518 (7537): 55–60. https://doi.org/10.1038/nature14121.

Marc Lipsitch [@mlipsitch]. 2022. 'Has BU Shown That They Are Capable of Self-Regulation, Assessing Both the Real Risks That Might Be Created and Documenting That They Considered Them and Found Them Minor Compared to Benefit: No. They Are in Full Denial Mode from Their Public Statements.' Tweet. *Twitter*. https://twitter.com/mlipsitch/status/1582582199370383362.

Marques, Andrew D., Michael Kummer, Oleksandr Kondratov, Arunava Banerjee, Oleksandr Moskalenko, and Sergei Zolotukhin. 2021. 'Applying Machine Learning to Predict Viral Assembly for Adeno-Associated Virus Capsid Libraries'. *Molecular Therapy - Methods & Clinical Development* 20 (March): 276–86. https://doi.org/10.1016/j.omtm.2020.11.017.

Martin, Vincent J. J., Douglas J. Pitera, Sydnor T. Withers, Jack D. Newman, and Jay D. Keasling. 2003. 'Engineering a Mevalonate Pathway in Escherichia Coli for Production of Terpenoids'. *Nature Biotechnology* 21 (7): 796–802. https://doi.org/10.1038/nbt833.

Maslanka, Susan E., Carolina Lúquez, Janet K. Dykes, William H. Tepp, Christina L. Pier, Sabine Pellett, Brian H. Raphael, et al. 2016. 'A Novel Botulinum Neurotoxin, Previously Reported as Serotype H, Has a Hybrid-Like Structure With Regions of Similarity to the Structures of Serotypes A and F and Is Neutralized With Serotype A Antitoxin'. *The Journal of Infectious Diseases* 213 (3): 379–85. https://doi.org/10.1093/infdis/jiv327.

Massung, R. F., L. I. Liu, J. Qi, J. C. Knight, T. E. Yuran, A. R. Kerlavage, J. M. Parsons, J. C. Venter, and J. J. Esposito. 1994. 'Analysis of the Complete Genome of Smallpox Variola Major Virus

Strain Bangladesh-1975'. *Virology* 201 (2): 215–40. https://doi.org/10.1006/viro.1994.1288.

Maurer, S. M., Keith V. Lucas, and S. N. Goldman. 2006. 'From Understanding to Action : Community-Based Options for Improving Security and Safety in Synthetic Biology'. *Undefined*. http://www.bioin.or.kr/InnoDS/data/upload/tech/Synthetic%20Biology.pdf.

McKenzie, Richard, Germar Bernhard, Ben Liley, Patrick Disterhoft, Steve Rhodes, Alkiviadis Bais, Olaf Morgenstern, et al. 2019. 'Success of Montreal Protocol Demonstrated by Comparing High-Quality UV Measurements with "World Avoided" Calculations from Two Chemistry-Climate Models'. *Scientific Reports* 9 (1): 12332. https://doi.org/10.1038/s41598-019-48625-z.

Meadows, Donella. 1999. 'Leverage Points: Places to Intervene in a System'. Sustainability Institute. http://www.donellameadows.org/wp-content/userfiles/Leverage_Points.pdf.

Meier, Anita F., Cornel Fraefel, and Michael Seyffert. 2020. 'The Interplay between Adeno-Associated Virus and Its Helper Viruses'. *Viruses* 12 (6). https://doi.org/10.3390/v12060662.

Meliani, Amine, Florence Boisgerault, Zachary Fitzpatrick, Solenne Marmier, Christian Leborgne, Fanny Collaud, Marcelo Simon Sola, et al. 2017. 'Enhanced Liver Gene Transfer and Evasion of Preexisting Humoral Immunity with Exosome-Enveloped AAV Vectors'. *Blood Advances* 1 (23): 2019–31. https://doi.org/10.1182/bloodadvances.2017010181.

Mikos, Georgios, Weitong Chen, and Junghae Suh. 2021. 'Machine Learning Identification of Capsid Mutations to Improve AAV Production Fitness'. *BioRxiv*, June, 2021.06.15.447941. https://doi.org/10.1101/2021.06.15.447941.

Miller, Judith, Stephen Engelberg, and William J. Broad. 2002. *Germs: Biological Weapons and America's Secret War*. Reprint edition. New York: Simon & Schuster.

Miller, Seumas. 2013. 'Moral Responsibility, Collective-Action Problems and the Dual-Use Dilemma in Science and Technology'. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 185–206. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.17.

Miller, Seumas, and Michael J. Selgelid. 2007. 'Ethical and Philosophical Consideration of the Dual-Use Dilemma in the Biological Sciences'. *Science and Engineering Ethics* 13 (4): 523–80. https://doi.org/10.1007/s11948-007-9043-4.

Millet, Jean K., Tiffany Tang, Lakshmi Nathan, Javier A. Jaimes, Hung-Lun Hsu, Susan Daniel, and Gary R. Whittaker. 2019. 'Production of Pseudotyped Particles to Study Highly Pathogenic Coronaviruses in a Biosafety Level 2 Setting'. *JoVE (Journal of Visualized Experiments)*, no. 145 (March): e59010. https://doi.org/10.3791/59010.

Millett, Piers, Thomas Binz, Sam Weiss Evans, Todd Kuiken, Ken Oye, Megan J. Palmer, Cécile van der Vlugt, Kathrina Yambao, and Samuel Yu. 2019. 'Developing a Comprehensive, Adaptive, and

International Biosafety and Biosecurity Program for Advanced Biotechnology: The IGEM Experience'. *Applied Biosafety* 24 (2): 64–71. https://doi.org/10.1177/1535676019838075.

Millett, Piers, and Paul Rutten. 2020. 'COVID-19, SARS-CoV-2, and Export Controls'. *Health Security* 18 (4): 329–34. https://doi.org/10.1089/hs.2020.0048.

Minehata, Masamichi, Judi Sture, Nariyoshi Shinomiya, and Simon Whitby. 2013. 'Implementing Biosecurity Education: Approaches, Resources and Programmes'. *Science and Engineering Ethics* 19 (4): 1473–86. https://doi.org/10.1007/s11948-011-9321-z.

Mitchell, Angela M., and R. Jude Samulski. 2013. 'Mechanistic Insights into the Enhancement of Adeno-Associated Virus Transduction by Proteasome Inhibitors'. *Journal of Virology* 87 (23): 13035–41. https://doi.org/10.1128/JVI.01826-13.

Mollentze, Nardus, Simon A. Babayan, and Daniel G. Streicker. 2021. 'Identifying and Prioritizing Potential Human-Infecting Viruses from Their Genome Sequences'. *PLOS Biology* 19 (9): e3001390. https://doi.org/10.1371/journal.pbio.3001390.

Monrad, Joshua T., and Rebecca Katz. 2020. 'Biosecurity, Biosafety, and the Management of Dangerous Pathogens for Public Health Research'. In *Viral Sovereignty and Technology Transfer: The Changing Global System for Sharing Pathogens for Public Health Research*, edited by Rebecca Katz and Sam F. Halabi, 100–119. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108676076.008.

Monrad, Joshua T., Jonas B. Sandbrink, and Neil G. Cherian. 2021. 'Promoting Versatile Vaccine Development for Emerging Pandemics'. *Npj Vaccines* 6 (1): 1–7. https://doi.org/10.1038/s41541-021-00290-y.

Msaouel, P., I. D. Iankov, A. Dispenzieri, and E. Galanis. 2012. 'Attenuated Oncolytic Measles Virus Strains as Cancer Therapeutics'. *Current Pharmaceutical Biotechnology* 13 (9): 1732–41. https://doi.org/10.2174/138920112800958896.

Mullin, Emily. 2017. 'Biohackers Disregard FDA Warning on DIY Gene Therapy'. *MIT Technology Review*, 1 December 2017. https://www.technologyreview.com/2017/12/01/147344/biohackers-disregard-fda-warning-on-diy-gene-therapy/.

Musunuri, Sriharshita, Jonas B. Sandbrink, Joshua T. Monrad, Megan J. Palmer, and Gregory D. Koblentz. 2021. 'Rapid Proliferation of Pandemic Research: Implications for Dual-Use Risks'. *MBio* 12 (5): e01864-21. https://doi.org/10.1128/mBio.01864-21.

National Academies of Sciences, Engineering, and Medicine. 1982. *Scientific Communication and National Security*. Washington, DC: National Academies Press. https://doi.org/10.17226/253.

———. 2016. *Gain-of-Function Research: Summary of the Second Symposium, March 10-11, 2016.*

Gain-of-Function Research: Summary of the Second Symposium, March 10-11, 2016. National Academies Press (US).

———. 2018. *Biodefense in the Age of Synthetic Biology*. Washington, DC: The National Academies Press. m.

National Human Genome Research Institute. 2012. 'Review of the Ethical, Legal and Social Implications Research Program and Related Activities (1990-1995)'. Genome.Gov. October 2012. https://www.genome.gov/10001747/elsi-program-review-19901995.

———. 2021. 'DNA Sequencing Costs: Data'. Genome.Gov. 2021. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

National Institute of Allergy and Infectious Diseases. 2020. 'Scoring & Summary Statements'. 19 August 2020. https://www.niaid.nih.gov/grants-contracts/scoring-summary-statements.

———. 2021. 'Review Criteria SOP'. 2021. https://www.niaid.nih.gov/research/review-criteria.

———. 2022. 'Submit an Application'. 2022. https://www.niaid.nih.gov/grants-contracts/submit-application.

National Institutes of Health. 2014a. 'Tools for the Identification, Assessment, Management, and Responsible Communication of Dual Use Research of Concern', September, 77.

———. 2014b. 'NIH Implementation of the US Government Policy on Institutional Oversight of Life Sciences Dual Use Research of Concern'. 21 November 2014. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-017.html.

———. 2019. 'Guidelines for Reviewers: Protections for Human'. March 2019. https://grants.nih.gov/grants/peer/guidelines_general/Guidelines_for_the_Review_of_the_Human_Subjects.pdf.

National Research Council. 2004. *Biotechnology Research in an Age of Terrorism*. Washington (DC): National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK222048/.

———. 2006. *Globalization, Biosecurity, and the Future of the Life Sciences*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11567.

———. 2009. *A Survey of Attitudes and Actions on Dual Use Research in the Life Sciences: A Collaborative Effort of the National Research Council and the American Association for the Advancement of Science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12460.

———. 2014. *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. National Academies Press.

National Research Council, Institute of Medicine, Board on Health Sciences Policy, Policy and Global Affairs, Committee on Science, Technology, and Law, Division on Earth and Life Studies, and Board on Life Sciences. 2015. *Potential Risks and Benefits of Gain-of-Function Research: Summary of a Workshop*. National Academies Press.

National Science Advisory Board for Biosecurity. 2007. 'Proposed Framework for the Oversight of Dual Use Life Sciences Research: Strategies for Minimizing the Potential Misuse of Research Information'. https://osp.od.nih.gov/wp-content/uploads/Proposed-Oversight-Framework-for-Dual-Use-Research.pdf.

———. 2016. 'Recommendations for the Evaluation and Oversight of Proposed Gain-of-Function Research'. https://osp.od.nih.gov/wp-content/uploads/2016/05/NSABB%20Working%20Group%20Draft%20Report%20-%20Recommendations%20for%20the%20Evaluation%20and%20Oversight%20of%20Proposed%20Gain-of-Function%20Research%205-6-2016.pdf.

———. 2023. 'Proposed Biosecurity Oversight Framework for the Future of Science'. https://osp.od.nih.gov/wp-content/uploads/2023/03/NSABB-Final-Report-Proposed-Biosecurity-Oversight-Framework-for-the-Future-of-Science.pdf.

National Security Decision Directives. 1985. 'NSDD-189: National Policy on the Transfer of Scientific, Technical and Engineering Information'. https://irp.fas.org/offdocs/nsdd/nsdd-189.htm.

Nationale Akademie der Wissenschaften Leopoldina. 2022. 'Gemeinsamer Ausschuss Dual Use 2'. 2022. https://www.leopoldina.org/ueber-uns/kooperationen/gemeinsamer-ausschuss-dual-use-2/dual-use/.

Nature Biotechnology. 2009. 'What's in a Name?' *Nature Biotechnology* 27 (12): 1071–73. https://doi.org/10.1038/nbt1209-1071.

Niu, Yuyu, Bin Shen, Yiqiang Cui, Yongchang Chen, Jianying Wang, Lei Wang, Yu Kang, et al. 2014. 'Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos'. *Cell* 156 (4): 836–43. https://doi.org/10.1016/j.cell.2014.01.027.

Normile, Dennis. 2021. 'Genetic Papers Containing Data from China's Ethnic Minorities Draw Fire'. *Science Insider*, August. https://doi.org/doi:10.1126/science.abl8764.

Noyce, Ryan S., and David H. Evans. 2018. 'Synthetic Horsepox Viruses and the Continuing Debate about Dual Use Research'. *PLOS Pathogens* 14 (10): e1007025. https://doi.org/10.1371/journal.ppat.1007025.

Noyce, Ryan S., Seth Lederman, and David H. Evans. 2018. 'Construction of an Infectious Horsepox Virus Vaccine from Chemically Synthesized DNA Fragments'. *PLOS ONE* 13 (1): e0188453. https://doi.org/10.1371/journal.pone.0188453.

Nuclear Threat Initiative and Johns Hopkins Center for Health Security. 2021. 'The 2021 Global Health Security Index'. https://www.ghsindex.org/.

Nuismer, Scott L. 2022. 'One Step Closer to a Transmissible Vaccine for Rabies Virus'. *PLOS Biology* 20 (4): e3001607. https://doi.org/10.1371/journal.pbio.3001607.

Nuismer, Scott L., and James J. Bull. 2020. 'Self-Disseminating Vaccines to Suppress Zoonoses'. *Nature Ecology & Evolution* 4 (July): 1168–73. https://doi.org/10.1038/s41559-020-1254-y.

O'Brien, John T., and Cassidy Nelson. 2020. 'Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology'. *Health Security* 18 (3): 219–27. https://doi.org/10.1089/hs.2019.0122.

Ofgem and Revealing Reality. 2020. 'Consumer Attitudes towards Decarbonisation and Net Zero'. https://www.ofgem.gov.uk/sites/default/files/docs/2020/10/consumer_attitudes_towards_decarbonisation_and_net_zero_1.pdf.

Ogden, Pierce J., Eric D. Kelsic, Sam Sinai, and George M. Church. 2019. 'Comprehensive AAV Capsid Fitness Landscape Reveals a Viral Gene and Enables Machine-Guided Design'. *Science* 366 (6469): 1139–43. https://doi.org/10.1126/science.aaw2900.

Open AI. 2020. 'OpenAI API'. OpenAI. 11 June 2020. https://openai.com/blog/openai-api/.

Open Philanthropy. 2022. 'Open Philanthropy Biosecurity Scholarships'. *Open Philanthropy* - (blog). 4 April 2022. https://www.openphilanthropy.org/open-philanthropy-biosecurity-scholarships/.

OpenAI. 2023. 'GPT-4 Technical Report'. arXiv. http://arxiv.org/abs/2303.08774.

Ott, Christian, and Jan Endrikat. 2022. 'Exploring the Association between Financial and Nonfinancial Carbon-Related Incentives and Carbon Performance'. *Accounting and Business Research* 0 (0): 1–34. https://doi.org/10.1080/00014788.2021.1993777.

Ouagrham-Gormley, Sonia Ben. 2014. *Barriers to Bioweapons: The Challenges of Expertise and Organization for Weapons Development*. 1st edition. Ithaca: Cornell University Press.

Our World in Data. 2022. 'Cumulative Confirmed COVID-19 Deaths by World Region'. Our World in Data. 2022. https://ourworldindata.org/grapher/cumulative-covid-deaths-region.

Owen, Anthony D. 2006. 'Renewable Energy: Externality Costs as Market Barriers'. *Energy Policy*, Hong Kong Editorial Board meeting presentations, 34 (5): 632–42. https://doi.org/10.1016/j.enpol.2005.11.017.

Oye, Kenneth A., Kevin M. Esvelt, Evan Appleton, Flaminia Catteruccia, George Church, Todd Kuiken, Shlomiya Bar-Yam Lightfoot, Julie McNamara, Andrea Smidler, and James P. Collins. 2014. 'Regulating Gene Drives'. *Science* 345 (6197): 626–28.

https://doi.org/10.1126/science.1254287.

Palmer, Megan J. 2020. 'Learning to Deal with Dual Use'. *Science* 367 (6482): 1057–1057. https://doi.org/10.1126/science.abb1466.

Palmer, Megan J., Francis Fukuyama, and David A. Relman. 2015. 'A More Systematic Approach to Biological Risk'. *Science* 350 (6267): 1471–73. https://doi.org/10.1126/science.aad8849.

Pandemic Preparedness Partnership. 2021. '100 Days Mission to Respond to Future Pandemic Threats'.

Pannu, Jaspreet, Jonas B. Sandbrink, Matthew Watson, Megan J. Palmer, and David A. Relman. 2021. 'Protocols and Risks: When Less Is More'. *Nature Protocols*, December, 1–2. https://doi.org/10.1038/s41596-021-00655-6.

Park, So Ra, and Jae Young Jang. 2021. 'The Impact of ESG Management on Investment Decision: Institutional Investors' Perceptions of Country-Specific ESG Criteria'. *International Journal of Financial Studies* 9 (3): 48. https://doi.org/10.3390/ijfs9030048.

Patel, Dylan, and Afzal Ahmad. 2023. 'Google "We Have No Moat, And Neither Does OpenAI"'. *Semianalysis* (blog). 4 May 2023. https://www.semianalysis.com/p/google-we-have-no-moat-and-neither.

Patlovich, Scott J., Robert J. Emery, Lawrence W. Whitehead, Eric L. Brown, and Rene Flores. 2015. 'Assessing the Biological Safety Profession's Evaluation and Control of Risks Associated with the Field Collection of Potentially Infectious Specimens'. *Applied Biosafety* 20 (1): 27–40. https://doi.org/10.1177/153567601502000104.

Patterson, Amy P., Lawrence A. Tabak, Anthony S. Fauci, Francis S. Collins, and Sally Howard. 2013. 'A Framework for Decisions About Research with HPAI H5N1 Viruses'. *Science* 339 (6123): 1036–37. https://doi.org/10.1126/science.1236194.

Perneger, Thomas V. 2005. 'The Swiss Cheese Model of Safety Incidents: Are There Holes in the Metaphor?' *BMC Health Services Research* 5 (November): 71. https://doi.org/10.1186/1472-6963-5-71.

Petro, James B., and David A. Relman. 2003. 'Understanding Threats to Scientific Openness'. *Science* 302 (5652): 1898–1898. https://doi.org/10.1126/science.1092493.

Pfeiffer, Julie K., and Terence S. Dermody. 2021. 'As Scientists Turn Their Attention to COVID-19, Other Research Is Not Getting Done – and That Can Have Lasting Consequences'. *The Conversation*, 29 January 2021. http://theconversation.com/as-scientists-turn-their-attention-to-covid-19-other-research-is-not-getting-done-and-that-can-have-lasting-consequences-154040.

Pichla-Gollon, Susan L., Shih-Wen Lin, Scott E. Hensley, Marcio O. Lasaro, Larissa Herkenhoff-Haut, Mark Drinker, Nia Tatsis, et al. 2009. 'Effect of Preexisting Immunity on an Adenovirus

Vaccine Vector: In Vitro Neutralization Assays Fail To Predict Inhibition by Antiviral Antibody In Vivo'. *Journal of Virology* 83 (11): 5567–73. https://doi.org/10.1128/JVI.00405-09.

Pickering, Jonathan, and Christian Barry. 2012. 'On the Concept of Climate Debt: Its Moral and Political Value'. *Critical Review of International Social and Political Philosophy* 15 (5): 667–85. https://doi.org/10.1080/13698230.2012.727311.

Pickett, Brett E., Douglas S. Greer, Yun Zhang, Lucy Stewart, Liwei Zhou, Guangyu Sun, Zhiping Gu, et al. 2012. 'Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for the Coronavirus Research Community'. *Viruses* 4 (11): 3209–26. https://doi.org/10.3390/v4113209.

Pigłowski, Marcin. 2021. 'The Intra-European Union Food Trade with the Relation to the Notifications in the Rapid Alert System for Food and Feed'. *International Journal of Environmental Research and Public Health* 18 (4): 1623. https://doi.org/10.3390/ijerph18041623.

Pine, Samuel O., James G. Kublin, Scott M. Hammer, Joleen Borgerding, Yunda Huang, Danilo R. Casimiro, and M. Juliana McElrath. 2011. 'Pre-Existing Adenovirus Immunity Modifies a Complex Mixed Th1 and Th2 Cytokine Response to an Ad5/HIV-1 Vaccine Candidate in Humans'. *PLoS ONE* 6 (4). https://doi.org/10.1371/journal.pone.0018526.

Pizzuto, Matteo Samuele, Micol Silic-Benussi, Vincenzo Ciminale, Ruth A. Elderfield, Ilaria Capua, and Wendy S.YR 2016 Barclay. 2016. 'An Engineered Avian-Origin Influenza A Virus for Pancreatic Ductal Adenocarcinoma Virotherapy'. *Journal of General Virology* 97 (9): 2166–79. https://doi.org/10.1099/jgv.0.000549.

Polack, Fernando P., Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, et al. 2020. 'Safety and Efficacy of the BNT162b2 MRNA Covid-19 Vaccine'. *New England Journal of Medicine* 383 (27): 2603–15. https://doi.org/10.1056/NEJMoa2034577.

Polanyi, M. 1974. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: ILUniversity of Chicago Press.

'Polis'. 2022. 2022. https://pol.is/home.

PREDICT Consortium. 2014. 'Reducing Pandemic Risk, Promoting Global Health'. One Health Institute, University of California, Davis. https://ohi.sf.ucdavis.edu/sites/g/files/dgvnsk5251/files/files/page/predict-final-report-lo.pdf.

PREDICT One Health Consortium. 2016. 'PREDICT Operating Procedures: Biosafety and Personal Protective Equipment (PPE) Use'.

Price, Jeffrey P. 2014. 'Effectiveness of Financial Incentives for Carbon Capture and Storage'. McLean,

Virginia, USA: Bluewave Resources. https://ieaghg.org/docs/General_Docs/Publications/Effectiveness%20of%20CCS%20Incentives.pdf.

Ramamoorth, Murali, and Aparna Narvekar. 2015. 'Non Viral Vectors in Gene Therapy- An Overview'. *Journal of Clinical and Diagnostic Research : JCDR* 9 (1): GE01–6. https://doi.org/10.7860/JCDR/2015/10443.5394.

Randerson, James. 2006. 'Revealed: The Lax Laws That Could Allow Assembly of Deadly Virus DNA'. The Guardian. 14 June 2006. http://www.theguardian.com/world/2006/jun/14/terrorism.topstories3.

Ranlet, Philip. 2000. 'The British, the Indians, and Smallpox: What Actually Happened at Fort Pitt in 1763?' *Pennsylvania History: A Journal of Mid-Atlantic Studies* 67 (3): 427–41.

Rappazzo, C. Garrett, Longping V. Tse, Chengzi I. Kaku, Daniel Wrapp, Mrunal Sakharkar, Deli Huang, Laura M. Deveau, et al. 2021. 'Broad and Potent Activity against SARS-like Viruses by an Engineered Human Monoclonal Antibody'. *Science* 371 (6531): 823–29. https://doi.org/10.1126/science.abf4830.

Rappert, Brian, and Michael J. Selgelid. 2013. *On the Dual Uses of Science and Ethics Principles, Practices, and Prospects*. ANU Press. https://doi.org/10.26530/OAPEN_462759.

Rauch, Susanne, Edith Jasny, Kim E. Schmidt, and Benjamin Petsch. 2018. 'New Vaccine Technologies to Combat Outbreak Situations'. *Frontiers in Immunology* 9: 1963. https://doi.org/10.3389/fimmu.2018.01963.

Razai, Mohammad S., Umar A. R. Chaudhry, Katja Doerholt, Linda Bauld, and Azeem Majeed. 2021. 'Covid-19 Vaccination Hesitancy'. *BMJ* 373 (May): n1138. https://doi.org/10.1136/bmj.n1138.

Rehman, Hasan, Ann W. Silk, Michael P. Kane, and Howard L. Kaufman. 2016. 'Into the Clinic: Talimogene Laherparepvec (T-VEC), a First-in-Class Intratumoral Oncolytic Viral Therapy'. *Journal for ImmunoTherapy of Cancer* 4 (1): 53. https://doi.org/10.1186/s40425-016-0158-5.

Rerks-Ngarm, Supachai, Punnee Pitisuttithum, Sorachai Nitayaphan, Jaranit Kaewkungwal, Joseph Chiu, Robert Paris, Nakorn Premsri, et al. 2009. 'Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand'. *New England Journal of Medicine* 361 (23): 2209–20. https://doi.org/10.1056/NEJMoa0908492.

Resnik, David B. 2013a. 'H5N1 Avian Flu Research and the Ethics of Knowledge'. *Hastings Center Report* 43 (2): 22–33. https://doi.org/10.1002/hast.143.

———. 2013b. 'Scientific Control Over Dual-Use Research: Prospects for Self-Regulation'. In . https://doi.org/10.22459/DUSE.12.2013.15.

Revill, James, and Catherine Jefferson. 2014. 'Tacit Knowledge and the Biological Weapons Regime'. *Science and Public Policy* 41 (5): 597–610. https://doi.org/10.1093/scipol/sct090.

Rip, Arie, and René Kemp. 1998. 'Technological Change'. *Human Choice and Climate Change: Vol. II, Resources and Technology*, 327–99.

Roberts, Diane M., Anjali Nanda, Menzo J.E. Havenga, Peter Abbink, Diana M. Lynch, Bonnie A. Ewald, Jinyan Liu, et al. 2006. 'Hexon-Chimaeric Adenovirus Serotype 5 Vectors Circumvent Pre-Existing Anti-Vector Immunity'. *Nature* 441 (7090): 239–43. https://doi.org/10.1038/nature04721.

Rosengard, Ariella M., Yu Liu, Zhiping Nie, and Robert Jimenez. 2002. 'Variola Virus Immune Evasion Design: Expression of a Highly Efficient Inhibitor of Human Complement'. *Proceedings of the National Academy of Sciences* 99 (13): 8808–13. https://doi.org/10.1073/pnas.112220499.

Rourke, Michelle F., Alexandra Phelan, and Charles Lawson. 2020. 'Access and Benefit-Sharing Following the Synthesis of Horsepox Virus'. *Nature Biotechnology* 38 (5): 537–39. https://doi.org/10.1038/s41587-020-0518-z.

Rovner, Alexis J., Adrian D. Haimovich, Spencer R. Katz, Zhe Li, Michael W. Grome, Brandon M. Gassaway, Miriam Amiram, et al. 2015. 'Recoded Organisms Engineered to Depend on Synthetic Amino Acids'. *Nature* 518 (7537): 89–93. https://doi.org/10.1038/nature14095.

Sadoff, Jerald, Glenda Gray, An Vandebosch, Vicky Cárdenas, Georgi Shukarev, Beatriz Grinsztejn, Paul A. Goepfert, et al. 2021. 'Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19'. *New England Journal of Medicine* 384 (23): 2187–2201. https://doi.org/10.1056/NEJMoa2101544.

Salloch, Sabine. 2018. 'The Dual Use of Research Ethics Committees: Why Professional Self-Governance Falls Short in Preserving Biosecurity'. *BMC Medical Ethics* 19 (1): 53. https://doi.org/10.1186/s12910-018-0295-0.

Salzberg, Steven. 2022. 'Gain-Of-Function Experiments At Boston University Create A Deadly New Covid-19 Virus. Who Thought This Was A Good Idea?' *Forbes*, 2022. https://www.forbes.com/sites/stevensalzberg/2022/10/24/gain-of-function-experiments-at-boston-university-create-a-deadly-new-covid-19-virus-who-thought-this-was-a-good-idea/.

Sandberg, Anders, and Cassidy Nelson. 2020. 'Who Should We Fear More: Biohackers, Disgruntled Postdocs, or Bad Governments? A Simple Risk Chain Model of Biorisk'. *Health Security* 18 (3): 155–63. https://doi.org/10.1089/hs.2019.0115.

Sandbrink, Jonas B., Janvi Ahuja, Jacob Swett, Gregory Koblentz, and Claire Standley. 2022. 'Mitigating Biosecurity Challenges of Wildlife Virus Discovery and Characterisation'. SSRN Scholarly Paper ID 4035760. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4035760.

Sandbrink, Jonas B., Ethan C. Alley, Matthew C. Watson, Gregory D. Koblentz, and Kevin M. Esvelt. 2022. 'Insidious Insights: Implications of Viral Vector Engineering for Pathogen Enhancement'. *Gene Therapy*, March. https://doi.org/10.1038/s41434-021-00312-3.

Sandbrink, Jonas B., Hamish Hobbs, Jacob Swett, Allan Dafoe, and Anders Sandberg. 2022. 'Differential Technology Development: A Responsible Innovation Principle for Navigating Technology Risks'. SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=4213670.

Sandbrink, Jonas B., and Gregory D. Koblentz. 2022. 'Biosecurity Risks Associated with Vaccine Platform Technologies'. *Vaccine* 40 (17): 2514–23. https://doi.org/10.1016/j.vaccine.2021.02.023.

Sandbrink, Jonas B., and Robin J. Shattock. 2020. 'RNA Vaccines: A Suitable Platform for Tackling Emerging Pandemics?' *Frontiers in Immunology* 11: 608460. https://doi.org/10.3389/fimmu.2020.608460.

Sandbrink, Jonas B., Matthew C. Watson, Andrew M. Hebbeler, and Kevin M. Esvelt. 2021. 'Safety and Security Concerns Regarding Transmissible Vaccines'. *Nature Ecology & Evolution* 5 (4): 405–6. https://doi.org/10.1038/s41559-021-01394-3.

Santiago-Ortiz, J., D. S. Ojala, O. Westesson, J. R. Weinstein, S. Y. Wong, A. Steinsapir, S. Kumar, I. Holmes, and D. V. Schaffer. 2015. 'AAV Ancestral Reconstruction Library Enables Selection of Broadly Infectious Viral Variants'. *Gene Therapy* 22 (12): 934–46. https://doi.org/10.1038/gt.2015.74.

Sarwar, Samreen, Sadaf Ilyas, Bilal Ahmed Khan, Danielle C. Lohman, Saleha Haffez, Madiha Rafique, Faizan Rashid, Junaid Akhtar, Saeed Khan, and Aurora O. Amoah. 2019. 'Awareness and Attitudes of Research Students Toward Dual-Use Research of Concern in Pakistan: A Cross-Sectional Questionnaire'. *Health Security* 17 (3): 229–39. https://doi.org/10.1089/hs.2019.0002.

Saxena, Manvendra, Thi Thu Hao Van, Fiona J. Baird, Peter J. Coloe, and Peter M. Smooker. 2013. 'Pre-Existing Immunity against Vaccine Vectors – Friend or Foe?' *Microbiology* 159 (Pt 1): 1–11. https://doi.org/10.1099/mic.0.049601-0.

Saylors, Karen, David J. Wolking, Emily Hagan, Stephanie Martinez, Leilani Francisco, Jason Euren, Sarah H. Olson, et al. 2021. 'Socializing One Health: An Innovative Strategy to Investigate Social and Behavioral Risks of Emerging Viral Threats'. *One Health Outlook* 3 (1): 11. https://doi.org/10.1186/s42522-021-00036-9.

Schönbrodt, Felix. 2016. 'Changing Hiring Practices towards Research Transparency: The First Open Science Statement in a Professorship Advertisement – Nicebread.De'. 6 January 2016. https://www.nicebread.de/open-science-hiring-practices/.

Schot, Johan, and Frank W. Geels. 2008. 'Strategic Niche Management and Sustainable Innovation Journeys: Theory, Findings, Research Agenda, and Policy'. *Technology Analysis & Strategic Management* 20 (5): 537–54. https://doi.org/10.1080/09537320802292651.

Schot, Johan, and Arie Rip. 1997. 'The Past and Future of Constructive Technology Assessment'. *Technological Forecasting and Social Change* 54 (2–3): 251–68. https://doi.org/10.1016/S0040-1625(96)00180-1.

Schulson, Michael. 2022. 'Experts Debate the Risks of Made-to-Order DNA'. *Undark Magazine*, 21 December 2022. https://undark.org/2022/12/21/experts-debate-the-risks-of-made-to-order-dna/.

Schweber, Silvan S. 2000. *In the Shadow of the Bomb: Oppenheimer, Bethe, and the Moral Responsibility of the Scientist*. *In the Shadow of the Bomb*. Princeton University Press. https://doi.org/10.1515/9781400849499.

Schwensow, Nina I, Brian Cooke, John Kovaliski, Ron Sinclair, David Peacock, Joerns Fickel, and Simone Sommer. 2014. 'Rabbit Haemorrhagic Disease: Virus Persistence and Adaptation in Australia'. *Evolutionary Applications* 7 (9): 1056–67. https://doi.org/10.1111/eva.12195.

Sebastian, Sarah, and Teresa Lambe. 2018. 'Clinical Advances in Viral-Vectored Influenza Vaccines'. *Vaccines* 6 (2): 29. https://doi.org/10.3390/vaccines6020029.

'Secure DNA Project'. 2022. 2022. https://www.securedna.org/main-en.

Selgelid, Michael J. 2013. 'Ethics and Dual-Use Research'. In *On the Dual Uses of Science and Ethics*, edited by Michael J. Selgelid and Brian Rappert, 3–10. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.6.

———. 2016. 'Gain-of-Function Research: Ethical Analysis'. *Science and Engineering Ethics* 22 (4): 923–64. https://doi.org/10.1007/s11948-016-9810-1.

Selgelid, Michael J., and Lorna Weir. 2010. 'The Mousepox Experience'. *EMBO Reports* 11 (1): 18–24. https://doi.org/10.1038/embor.2009.270.

Shahryari, Alireza, Marie Saghaeian Jazi, Saeed Mohammadi, Hadi Razavi Nikoo, Zahra Nazari, Elaheh Sadat Hosseini, Ingo Burtscher, Seyed Javad Mowla, and Heiko Lickert. 2019. 'Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders'. *Frontiers in Genetics* 10. https://doi.org/10.3389/fgene.2019.00868.

Shao, Wenwei, Xiaojing Chen, Richard J Samulski, Matthew L Hirsch, and Chengwen Li. 2018. 'Inhibition of Antigen Presentation during AAV Gene Therapy Using Virus Peptides'. *Human Molecular Genetics* 27 (4): 601–13. https://doi.org/10.1093/hmg/ddx427.

Shavell, Steven. 1984. 'Liability for Harm versus Regulation of Safety'. *The Journal of Legal Studies* 13

(2): 357–74.

Sheahan, Timothy P., Amy C. Sims, Rachel L. Graham, Vineet D. Menachery, Lisa E. Gralinski, James B. Case, Sarah R. Leist, et al. 2017. 'Broad-Spectrum Antiviral GS-5734 Inhibits Both Epidemic and Zoonotic Coronaviruses'. *Science Translational Medicine* 9 (396): eaal3653. https://doi.org/10.1126/scitranslmed.aal3653.

Sheahan, Timothy P., Amy C. Sims, Shuntai Zhou, Rachel L. Graham, Andrea J. Pruijssers, Maria L. Agostini, Sarah R. Leist, et al. 2020. 'An Orally Bioavailable Broad-Spectrum Antiviral Inhibits SARS-CoV-2 in Human Airway Epithelial Cell Cultures and Multiple Coronaviruses in Mice'. *Science Translational Medicine* 12 (541): eabb5883. https://doi.org/10.1126/scitranslmed.abb5883.

Shevlane, Toby. 2022. 'Structured Access: An Emerging Paradigm for Safe AI Deployment'. arXiv. http://arxiv.org/abs/2201.05159.

Shevlane, Toby, Sebastian Farquhar, Ben Garfinkel, Mary Phuong, Jess Whittlestone, Jade Leung, Daniel Kokotajlo, et al. 2023. 'Model Evaluation for Extreme Risks'. arXiv. https://doi.org/10.48550/arXiv.2305.15324.

Sinai, Sam, Nina Jain, George M. Church, and Eric D. Kelsic. 2021. 'Generative AAV Capsid Diversification by Latent Interpolation'. *BioRxiv*, April, 2021.04.16.440236. https://doi.org/10.1101/2021.04.16.440236.

Singh, Ravi, Bowen Tian, and Kostas Kostarelos. 2008. 'Artificial Envelopment of Nonenveloped Viruses: Enhancing Adenovirus Tumor Targeting in Vivo'. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 22 (9): 3389–3402. https://doi.org/10.1096/fj.08-103275.

Skalmeraas, Olaf. 2017. 'Sleipner Carbon Capture and Storage Project'. Institution of Civil Engineers (ICE). February 2017. https://www.ice.org.uk/engineering-resources/case-studies/sleipner-carbon-capture-and-storage-project/.

Smith, James Andrew, and Jonas B. Sandbrink. 2022. 'Biosecurity in an Age of Open Science'. *PLOS Biology* 20 (4): e3001600. https://doi.org/10.1371/journal.pbio.3001600.

Soice, Emily H., Rafael Rocha, Kimberlee Cordova, Michael Specter, and Kevin M. Esvelt. 2023. 'Can Large Language Models Democratize Access to Dual-Use Biotechnology?' arXiv. https://doi.org/10.48550/arXiv.2306.03809.

Solomon, Aris, and Linda Lewis. 2002. 'Incentives and Disincentives for Corporate Environmental Disclosure'. *Business Strategy and the Environment* 11 (3): 154–69. https://doi.org/10.1002/bse.328.

Specter, Michael. 2016. 'Rewriting the Code of Life'. *The New Yorker*, 25 December 2016.

https://www.newyorker.com/magazine/2017/01/02/rewriting-the-code-of-life.

Starr, Tyler N., Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H. D. Crawford, Adam S. Dingens, Mary Jane Navarro, et al. 2020. 'Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding'. *Cell* 182 (5): 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012.

Steffen, Imke, Nathan M. Liss, Bradley S. Schneider, Joseph N. Fair, Charles Y. Chiu, and Graham Simmons. 2013. 'Characterization of the Bas-Congo Virus Glycoprotein and Its Function in Pseudotyped Viruses'. *Journal of Virology* 87 (17): 9558–68. https://doi.org/10.1128/JVI.01183-13.

Steinbruner, John, Elisa D. Harris, Nancy Gallagher, and Stacy M. Okutani. 2007. 'Controlling Dangerous Pathogens: A Prototype Protective Oversight System'. The Center for International and Security Studies at Maryland. https://drum.lib.umd.edu/bitstream/handle/1903/7949/pathogens_project_monograph.pdf?sequence=1&isAllowed=y.

Stilgoe, Jack, Richard Owen, and Phil Macnaghten. 2013. 'Developing a Framework for Responsible Innovation'. *Research Policy* 42 (9): 1568–80. https://doi.org/10.1016/j.respol.2013.05.008.

Strassburg, M. A. 1982. 'The Global Eradication of Smallpox'. *American Journal of Infection Control* 10 (2): 53–59. https://doi.org/10.1016/0196-6553(82)90003-7.

Sture, Judi. 2013. 'Moral Development and Ethical Decision-Making'. In *On the Dual Uses of Science and Ethics*, edited by Brian Rappert and Michael J. Selgelid, 97–120. Principles, Practices, and Prospects. ANU Press. https://www.jstor.org/stable/j.ctt5hgz15.12.

Sundaram, Lalitha S. 2021. 'Biosafety in DIY-Bio Laboratories: From Hype to Policy'. *EMBO Reports* 22 (4): e52506. https://doi.org/10.15252/embr.202152506.

Sunstein, Cass R. 2005. *Laws of Fear: Beyond the Precautionary Principle*. The Seeley Lectures. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790850.

Szybalski, W. 1974. 'In Vivo and in Vitro Initiation of Transcription'. In *Control of Gene Expression*, edited by Alexander Kohn and Adam Shatkay, 23–24. Advances in Experimental Medicine and Biology. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-3246-6_3.

Ta, Lisa, Laura Gosa, and David A. Nathanson. 2018. 'Biosafety and Biohazards: Understanding Biosafety Levels and Meeting Safety Requirements of a Biobank'. *Biobanking* 1897 (December): 213–25. https://doi.org/10.1007/978-1-4939-8935-5_19.

Taubenberger, Jeffery K., Ann H. Reid, Raina M. Lourens, Ruixue Wang, Guozhong Jin, and Thomas G. Fanning. 2005. 'Characterization of the 1918 Influenza Virus Polymerase Genes'. *Nature* 437 (7060): 889–93. https://doi.org/10.1038/nature04230.

Thacker, Erin E, Laura Timares, and Qiana L Matthews. 2009. 'Strategies to Overcome Host Immunity to Adenovirus Vectors in Vaccine Development'. *Expert Review of Vaccines* 8 (6): 761–77. https://doi.org/10.1586/erv.09.29.

Thadani, Nicole N., Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Chris Sander, Yarin Gal, and Debora S. Marks. 2023. 'Learning from Pre-Pandemic Data to Forecast Viral Escape'. bioRxiv. https://doi.org/10.1101/2022.07.21.501023.

The Australia Group. 2020. 'List of Human and Animal Pathogens and Toxins for Export Control'. 2020. https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/human_animal_pathogens.html.

The Brink Staff. 2022. 'NEIDL Researchers Refute UK Article about COVID Strain'. *The Brink* (blog). 2022. https://www.bu.edu/articles/2022/neidl-researchers-refute-uk-article-about-covid-strain/.

The White House. 2021. 'American Pandemic Preparedness: Transforming Our Capabilities'. https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf?page=29.

———. 2022. 'Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy'. The White House. 12 September 2022. https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/.

Thi Nhu Thao, Tran, Fabien Labroussaa, Nadine Ebert, Philip V'kovski, Hanspeter Stalder, Jasmine Portmann, Jenna Kelly, et al. 2020. 'Rapid Reconstruction of SARS-CoV-2 Using a Synthetic Genomics Platform'. *Nature* 582 (7813): 561–65. https://doi.org/10.1038/s41586-020-2294-9.

Thomas, Ann. 1970. *Legal Limits on the Use of Chemical and Biological Weapons*. Southern Methodist Univ Pr.

Torre, Daniel de la, and Jason W. Chin. 2021. 'Reprogramming the Genetic Code'. *Nature Reviews. Genetics* 22 (3): 169–84. https://doi.org/10.1038/s41576-020-00307-7.

Torres, Phil. 2018. 'Who Would Destroy the World? Omnicidal Agents and Related Phenomena'. *Aggression and Violent Behavior* 39 (March): 129–38. https://doi.org/10.1016/j.avb.2018.02.002.

Tucker, Jonathan B., ed. 2012. *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies*. The MIT Press. https://www.jstor.org/stable/j.ctt5vjq1n.

Tucker, Jonathan B., and Eric R. Mahan. 2009. 'President Nixon's Decision to Renounce the U.S. Offensive Biological Weapons Program'. Center for the Study of Weapons of Mass Destruction Case Study 1. Washington, DC: National Defense University Press. https://ndupress.ndu.edu/Publications/Article/718029/president-nixons-decision-to-renounce-the-us-offensive-biological-weapons-progr/https%3A%2F%2Fndupress.ndu.edu%2FMedia%2FNews%2FNews-Article-View%2FArticle%2F718029%2Fpresident-nixons-decision-to-renounce-the-us-offensive-biological-weapons-progr%2F.

Tumpey, Terrence M., Christopher F. Basler, Patricia V. Aguilar, Hui Zeng, Alicia Solórzano, David E. Swayne, Nancy J. Cox, et al. 2005. 'Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus'. *Science (New York, N.Y.)* 310 (5745): 77–80. https://doi.org/10.1126/science.1119392.

Twist. 2023. 'High Quality Gene Synthesis - Twist Bioscience'. 2023. https://www.twistbioscience.com/products/genes.

U. S. Atomic Energy Commission. 1954. *In the Matter of J. Robert Oppenheimer: Transcript of Hearing before Personnel Security Board, Washington, D.C., April 12, 1954, through May 5, 1954*. Washington, DC: U.S. Government Printing Office.

UK Government. 2021. 'COP26 Declaration on Accelerating the Transition to 100% Zero Emission Cars and Vans'. https://www.gov.uk/government/publications/cop26-declaration-zero-emission-cars-and-vans/cop26-declaration-on-accelerating-the-transition-to-100-zero-emission-cars-and-vans.

Ulrich Dirnagl. 2018. 'If You Are Applying for a Professorship at the Charite You Now Need to Tell Us about Your Contributions to Your Scientific Field, Open Science, Team Science, Interactions with Stakeholders. Past and Future Plans. As a Structured Narrative. Https://T.Co/Lm3aXBGSE0'. Tweet. *Twitter*. https://twitter.com/dirnagl/status/970227847943114752.

United Nations Office for Disarmament Affairs. 2022. *UN Security Council Resolution 1540 (2004)*. https://www.un.org/disarmament/wmd/sc1540/.

United States Senate Committee on Environment and Public Works. 2002. 'Price-Anderson Act Reauthorization Hearing', January.

UNODA. 2022. 'Biological Weapons Convention'. 2022. https://www.un.org/disarmament/biological-weapons/.

Ura, Takehiro, Kenji Okuda, and Masaru Shimada. 2014. 'Developments in Viral Vector-Based Vaccines'. *Vaccines* 2 (3): 624–41. https://doi.org/10.3390/vaccines2030624.

Urbina, Fabio, Filippa Lentzos, Cédric Invernizzi, and Sean Ekins. 2022. 'Dual Use of Artificial-

Intelligence-Powered Drug Discovery'. *Nature Machine Intelligence* 4 (3): 189–91. https://doi.org/10.1038/s42256-022-00465-9.

U.S. Department of Health and Human Services. 2010. 'Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA'. https://aspr.hhs.gov:443/legal/syndna/Pages/default.aspx.

———. 2012. 'United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern'. https://www.phe.gov/s3/dualuse/Documents/us-policy-durc-032812.pdf.

———. 2014. 'United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern'. https://www.phe.gov/s3/dualuse/documents/oversight-durc.pdf.

———. 2017. 'Framework for Guiding Funding Decisions about Proposed Research Involving Enhanced Potential Pandemic Pathogens'. https://www.phe.gov/s3/dualuse/Documents/P3CO.pdf.

———. 2021. 'Research Instructions for NIH and Other PHS Agencies'.

USAID. 2021. 'USAID Announces New $125 Million Project To Detect Unknown Viruses With Pandemic Potential'. 5 October 2021. https://www.usaid.gov/news-information/press-releases/oct-5-2021-usaid-announces-new-125-million-project-detect-unknown-viruses.

Vickers, Claudia E., and Paul S. Freemont. 2022. 'Pandemic Preparedness: Synthetic Biology and Publicly Funded Biofoundries Can Rapidly Accelerate Response Time'. *Nature Communications* 13 (1): 453. https://doi.org/10.1038/s41467-022-28103-3.

Vinke, Svenja, Irina Rais, and Piers Millett. 2022. 'The Dual-Use Education Gap: Awareness and Education of Life Science Researchers on Nonpathogen-Related Dual-Use Research'. *Health Security* 20 (1): 35–42. https://doi.org/10.1089/hs.2021.0177.

Vogel, Kathleen M. 2012. *Phantom Menace or Looming Danger? A New Framework for Assessing Bioweapons Threats*. Baltimore, MD: Johns Hopkins University Press.

Wang, Lian, Yonggang Zhang, Dongguang Wang, Xiang Tong, Tao Liu, Shijie Zhang, Jizhen Huang, et al. 2021. 'Artificial Intelligence for COVID-19: A Systematic Review'. *Frontiers in Medicine* 8. https://www.frontiersin.org/articles/10.3389/fmed.2021.704256.

Warrick, Jo. 2006. 'Suspect and A Setback In Al-Qaeda Anthrax Case <span Class='. *Washington Post*, 31 October 2006. https://www.washingtonpost.com/archive/politics/2006/10/31/suspect-and-a-setback-in-al-qaeda-anthrax-case-span-classbankheadscientist-with-ties-to-group-goes-freespan/eeb4e5a1-9d08-4dfa-bccc-5c18e311502a/.

Watson, Joseph L., David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, et al. 2022. 'Broadly Applicable and Accurate Protein Design by

Integrating Structure Prediction Networks and Diffusion Generative Models'. bioRxiv. https://doi.org/10.1101/2022.12.09.519842.

Weaver, Eric A., and Michael A. Barry. 2008. 'Effects of Shielding Adenoviral Vectors with Polyethylene Glycol on Vector-Specific and Vaccine-Mediated Immune Responses'. *Human Gene Therapy* 19 (12): 1369–82. https://doi.org/10.1089/hum.2008.091.

Wec, Anna Z., Kathy S. Lin, Jamie C. Kwasnieski, Sam Sinai, Jeff Gerold, and Eric D. Kelsic. 2021. 'Overcoming Immunological Challenges Limiting Capsid-Mediated Gene Therapy With Machine Learning'. *Frontiers in Immunology* 12: 1443. https://doi.org/10.3389/fimmu.2021.674021.

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, et al. 2022. 'Emergent Abilities of Large Language Models'. arXiv. https://doi.org/10.48550/arXiv.2206.07682.

Westfall, Patrick J., Douglas J. Pitera, Jacob R. Lenihan, Diana Eng, Frank X. Woolard, Rika Regentin, Tizita Horning, et al. 2012. 'Production of Amorphadiene in Yeast, and Its Conversion to Dihydroartemisinic Acid, Precursor to the Antimalarial Agent Artemisinin'. *Proceedings of the National Academy of Sciences* 109 (3): E111–18. https://doi.org/10.1073/pnas.1110740109.

Wheelis, Mark. 1998. 'First Shots Fired in Biological Warfare'. *Nature* 395 (6699): 213–213. https://doi.org/10.1038/26089.

Wille, Michelle, Jemma L. Geoghegan, and Edward C. Holmes. 2021. 'How Accurately Can We Assess Zoonotic Risk?' *PLOS Biology* 19 (4): e3001135. https://doi.org/10.1371/journal.pbio.3001135.

Wilton Park. 2013. 'Dual-Use Biology: How to Balance Open Science with Security (WP1260)'. https://www.wiltonpark.org.uk/event/wp1260/.

Wintle, Bonnie C, Christian R Boehm, Catherine Rhodes, Jennifer C Molloy, Piers Millett, Laura Adam, Rainer Breitling, et al. 2017. 'A Transatlantic Perspective on 20 Emerging Issues in Biological Engineering'. Edited by Peter A Rodgers. *ELife* 6 (November): e30247. https://doi.org/10.7554/eLife.30247.

Woo, Gordon. 2002. 'Quantitative Terrorism Risk Assessment'. *The Journal of Risk Finance* 4 (1): 7–14. https://doi.org/10.1108/eb022949.

World Economic Forum and Nuclear Threat Initiative. 2020. 'Biosecurity Innovation and Risk Reduction: A Global Framework for Accessible, Safe and Secure DNA Synthesis'. Geneva. https://media.nti.org/documents/Biosecurity_Innovation_and_Risk_Reduction.pdf.

World Health Assembly. 2002. 'Global Public Health Response to Natural Occurrence, Accidental Release or Deliberate Use of Biological and Chemical Agents or Radionuclear Material That

Affect Health'. https://apps.who.int/gb/archive/pdf_files/WHA55/ewha5516.pdf.

World Health Organisation. 2010. 'Responsible Life Sciences Research for Global Health Security'. 2010. https://web.archive.org/web/20141020125107/http://www.who.int/csr/resources/publications/HSE_GAR_BDP_2010_2/en/.

———. 2012. 'Report on Technical Consultation on H5N1 Research Issues'. Geneva. https://www.who.int/publications/m/item/report-on-technical-consultation-on-h5n1-research-issues.

———. 2013. 'Report of the WHO Informal Consultation on Dual Use Research of Concern'. Geneva. https://www.who.int/csr/durc/durc_feb2013_full_mtg_report.pdf.

———. 2020. 'Laboratory Biosafety Manual, 4th Edition'. Geneva, Switzerland. https://www.who.int/publications-detail-redirect/9789240011311.

———. 2021a. '26 International Experts to Kickstart the One Health High Level Expert Panel (OHHLEP)'. 2021. https://www.who.int/news/item/11-06-2021-26-international-experts-to-kickstart-the-joint-fao-oie-unep-who-one-health-high-level-expert-panel-(ohhlep).

———. 2021b. 'Emerging Technologies and Dual-Use Concerns: A Horizon Scan for Global Public Health'. Geneva. https://www.who.int/publications-detail-redirect/9789240036161.

———. 2021c. 'Dual Use Life Science Research (DUR/C): Dialogue with Academies and Councils: Meeting Report, 6 July 2020'. Geneva. https://www.who.int/publications-detail-redirect/9789240031999.

———. 2022a. 'Joint External Evaluation Tool: International Health Regulations (2005) - Third Edition'. https://www.who.int/publications-detail-redirect/9789240051980.

———. 2022b. 'Global Guidance Framework for the Responsible Use of the Life Sciences: Mitigating Biorisks and Governing Dual-Use Research'. Geneva, Switzerland. https://www.who.int/publications-detail-redirect/9789240056107.

Wright, Lawrence. 2002. 'The Man Behind Bin Laden'. *The New Yorker*, 8 September 2002. https://www.newyorker.com/magazine/2002/09/16/the-man-behind-bin-laden.

Xie, Xuping, Kumari G. Lokugamage, Xianwen Zhang, Michelle N. Vu, Antonio E. Muruato, Vineet D. Menachery, and Pei-Yong Shi. 2021. 'Engineering SARS-CoV-2 Using a Reverse Genetic System'. *Nature Protocols*, January, 1–24. https://doi.org/10.1038/s41596-021-00491-8.

Xin, Ke-Qin, Hiroaki Mizukami, Masashi Urabe, Yoshihiko Toda, Kaori Shinoda, Atsushi Yoshida, Kenji Oomura, et al. 2006. 'Induction of Robust Immune Responses against Human Immunodeficiency Virus Is Supported by the Inherent Tropism of Adeno-Associated Virus

Type 5 for Dendritic Cells'. *Journal of Virology* 80 (24): 11899–910. https://doi.org/10.1128/JVI.00890-06.

Xue, Yang, Lijun Shang, and Weiwen Zhang. 2021. 'Building and Implementing a Multi-Level System of Ethical Code for Biologists under the Biological and Toxin Weapons Convention (BTWC) of the United Nations'. *Journal of Biosafety and Biosecurity* 3 (2): 108–19. https://doi.org/10.1016/j.jobb.2021.09.001.

Yang, Jingfeng, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. 'Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond'. arXiv. http://arxiv.org/abs/2304.13712.

Yao, Tianzhuo, Xueying Zhou, Chuanling Zhang, Xiaojuan Yu, Zhenyu Tian, Lihe Zhang, and Demin Zhou. 2017. 'Site-Specific PEGylated Adeno-Associated Viruses with Increased Serum Stability and Reduced Immunogenicity'. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry* 22 (7): 1155. https://doi.org/10.3390/molecules22071155.

Zhang, Zheng, Zena Cai, Zhiying Tan, Congyu Lu, Taijiao Jiang, Gaihua Zhang, and Yousong Peng. 2019. 'Rapid Identification of Human-Infecting Viruses'. *Transboundary and Emerging Diseases* 66 (6): 2517–22. https://doi.org/10.1111/tbed.13314.

Zimmer, Carl, and James Gorman. 2021. 'Fight Over Covid's Origins Renews Debate on Risks of Lab Work'. *The New York Times*, 20 June 2021, sec. Science. https://www.nytimes.com/2021/06/20/science/covid-lab-leak-wuhan.html.