Balancing Broader Impacts:

A Study of the NSF's Merit Review Criteria from Individual and Institutional Perspectives

by

Dania Noelle Wright

A Dissertation
Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2024 by the
Graduate Supervisory Committee:

Andrew Maynard, Chair
Shawn Jordan
Ruth Wylie

ARIZONA STATE UNIVERSITY

December 2024

ABSTRACT

This dissertation examines how the National Science Foundation's (NSF) Merit Review criteria, particularly Broader Impacts (BI), are perceived and valued by researchers, former Program Officers, and institutional officials. Through analysis of surveys, semi-structured interviews, and Large Language Model-assisted transcription of public discussions from the 2023-2024 National Science Board – National Science Foundation Task Force on Merit Review, this study reveals that Intellectual Merit (IM) criterion continues to be prioritized by all parties over Broader Impacts in both the proposal review process and career advancement. Despite increasing support for Broader Impacts, concerns persist about the accessibility of institutional support and the adequacy of default Broader Impacts activities, such as broadening participation through training underrepresented students, which are often seen as superficial. Task Force discussions further highlight how institutional policies shape the official interpretation and implementation of these criteria, emphasizing the evolving concerns over weighting, implementation, and accountability at institutional levels. The findings suggest that the NSF and academic institutions must enhance guidance and support for integrating both criteria, advocating for more equitable recognition of efforts related to Broader Impacts. The development of "Impact Identities" and the establishment of Broader Impacts centers are proposed as strategies to embed societal impact into research practices. However, such centers must be carefully managed to avoid prioritizing fundability over meaningful researcher involvement. Future research should include community and industry perspectives to better understand the broader reception and value of NSF-funded research beyond the academic community.

DEDICATION

To my loving husband and wonderful family, for supporting and encouraging me even though my degree is impossible to remember and you thought I was researching something that sounded like a bizarre sci-fi plot. Thank you for believing in me when I didn't believe in myself.

And to my dogs, Luna and Percival, who helped me remember sometimes all you need is a good nap to make the world make sense again.

ACKNOWLEDGMENTS

This dissertation would not exist without the support, encouragement, and honest feedback of countless people—more than I can ever list here, but I will do my best to try. First and foremost, I express my deepest gratitude to my dissertation committee chair, Dr. Andrew Maynard. Your guidance, insight, and unwavering support—especially through my many topic pivots—have been invaluable. Your mentorship in exploring the intersections of science and society has profoundly shaped my work. I am truly fortunate to have had the opportunity to learn from and work with you.

I am equally grateful to Dr. Shawn Jordan, who not only served on my committee but also welcomed me into his STEAM lab, dedicating countless hours to helping me grow as a researcher, writer, and learner. Your guidance has been incredibly valuable, and I'm thankful you took a chance on another interdisciplinary oddball like me. My sincere thanks also go to Dr. Ruth Wylie, whose unique perspectives and quiet support and guidance as a committee member have been crucial to my dissertation's development and I'm so glad you agreed to be on my committee!

Beyond my committee, I've been so lucky to find amazing mentorship and community throughout this process. Dr. David Guston's deep institutional knowledge has been invaluable since I began this PhD, shaping my research approach in profound ways. I'm also grateful to Dr. Darshan Karwat, whose example as a teacher and scholar has significantly influenced my research trajectory. To my HSD cohort—your support, advice, and memes have made this journey much more enjoyable, and we'll need to find a chance to celebrate at Postino's after all is said and done!

I would also like to extend heartfelt thanks to the members of various writing groups, accountability groups, and research labs I've been a part of, beginning from my first day in the program with the wonderful members of the now-defunct Pudding Club—Nicole Mayberry, Bobby Sickler, Jason Brown, and Jeishu Wang—who helped me navigate graduate school during a global pandemic. Particular thanks also go to the STEAM Lab's WHAM (Writing Hebdomadally & Accountability Meeting) sessions, and especially to Steven Weiner, Cecelia La Place, and Ieshya Anderson for your valuable feedback and time. I am also incredibly thankful to my Dissertation Writing Group peers—Kristina Vera Philips, Emily Singer, and Elizabeth Williams—whose twice-weekly writing sessions and unwavering support helped me reach this point.

I am perpetually grateful for my family—my mother and father, Ofelia and William Wright, who gave me the tools to succeed, and my sister, Dr. Chelsea Wright, who helped me navigate the PhD journey in ways only someone who's been through it can. And of course, to my friends who've stuck by me through this absurd challenge—thank you. From old college friends in Book Club to my longsword students and friends at the Phoenix Society of Historical Swordsmanship, your cheerleading and patience has been so appreciated.

Lastly, I owe the biggest thanks of all to my amazing husband, James Grant, whose unwavering love and unflagging belief in me has made all the difference. This dissertation would not exist without your support, and there aren't enough words to express how grateful I am to have you in my life. I love you and our life together (and our dogs) more than anything and I'm in awe every day that I get to explore the future with you. I look forward to seeing what life after the doctorate looks like!

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

This dissertation investigates the processes through which the National Science Foundation's (NSF) merit review criteria are interpreted and applied by various stakeholders both outside and inside the Foundation. The research therein considers both the personal perspectives of individual researchers and former Program Officers, as well as the perspectives of institutional officials within both the NSF and its governing board, the National Science Board, with the goal of providing a detailed analysis of their understanding and implementation of these criteria.

This study employs a mixed methods approach, integrating survey data, semi-structured interview responses, and public meeting notes from a joint National Science Board and National Science Foundation Task Force on the Merit Review criteria. These data sources are brought together to illustrate a dynamic range of expectations, needs, and tensions in the proposal submission and selection process within one of the largest administrators of scientific funding both in the United States and globally. Given the National Science Foundation's influential role in shaping research priotities across a wide range of institutions nationwide, understanding how the current Merit Review criteria influence the distribution of scientific grants is critical for both researchers and institutions across countless fields and disciplines.

1. Historical Overview

Various countries have implemented different approaches to directing and quantifying research impact as a part of larger research bodies, as seen with the United Kingdom's Research Excellence Framework (REF) as implemented by the United Kingdom Research and Innovation (UKRI) body (whose definition for research impact is noted above), as well as Canada's Social Science and Humanities Research Council (SSHRC), and the United States' National Science Foundation

(Davis and Laas, 2013). From this breadth, various research universities and other research institutions have interpreted those frameworks internally with various results. These numerous frameworks and interpretations may obscure the foundational questions around the role of these various organizing and funding bodies in defining and delimiting what qualifies as 'impact', particularly in contrast to understandings of and priorities around impact among individual researchers themselves. Directly entering this fray is the National Science Foundation with the continued use of the "Broader Impacts" criterion as a means of assessing the merit of proposed research projects.

Amongst its global peers, the National Science Foundation is among the oldest and most highly funded research bodies in the world, with an annual operating budget around 8.28 billion US dollars in fiscal year 2020 and 7.7 billion US dollars in fiscal year 2021, and expected to engage over 59,000 researchers and scholars and an estimated 325,000 individuals directly in 2023, including college, graduate, and K-12 students (National Science Foundation, 2024). While the European Research Council (ERC) surpasses the NSF in both budget and reach (European Research Council, 2023), the single-nation focus of the NSF may minimize cross- and inter-cultural variation and bias in approaches to the role of science and research impact. As such, this proposal focuses on just one research body and its approach to both science and its potential impact: the National Science Foundation.

The National Science Foundation (NSF) was established in 1950 in the aftermath of World War II. Its primary goal was to promote the progress of science, advance national health, prosperity, and welfare, and secure national defense, among other purposes. The foundation emerged in response to the growing need to fund scientific research and development independently from military or other governmental projects, marking a pivotal moment in the development of science policy in the United States.

The evolution of NSF's merit review criteria has been significant, beginning with the establishment of the organization in 1950 with a single criterion, "the scientific merit of the proposed research, including the competence of the investigator" (Mazuzan, 1992). The following year, with the first influx of proposals, that single criterion was quickly expanded for practical reasons to four primary criteria for assessing the value and fundability of research: (1) researcher performance competence, (2) the intrinsic merit of the research, (3) the utility or relevance of the research, and (4) its effect on the infrastructure of science and engineering (Rothenberg, 2010).

In 1996, after several years of internal discussion (Rothenberg, 2010; Mazuzan, 1992), the NSF committed to a reassessment of Merit Review process, convening the NSB-NSF Staff Merit Review Task Force to address the question of reimagining the proposal selection and award process. After both significant internal deliberation and a required period of public comment, and following the recommendations of that first Task Force, in 1997 the NSF instituted two official Merit Review criteria: Intellectual Merit (IM) and Broader Impacts (BI) (NSB-NSF Staff Merit Review Task Force, 1997). The revised criteria, along with guiding questions as recommended by the 1997 Task Force, are shown in Table 1.

This bifurcation was designed to ensure that funded research would contribute both to scientific advancement and to broader societal goals, primarily driven by growing calls for accountability from Congress and the federal government. The 1997 change in criteria, and particularly the creation of the Broader Impacts criterion, has been attributed to a increased focus on the NSF's national impact as a result of the Science and Engineering Equal Opportunities Act of 1980 (Chubin, 2007) and the Government Performance Results Act of 1993 (Holbrook, 2012). These efforts focused attention on results and tied the idea of external outcomes to internal funding, driving efforts to raise the profile of 'socially beneficial' research.

Table 1. Format of the Merit Review criteria as proposed in 1997.

Criterion	**Guiding Questions**
Intellectual Merit:	*"The following are suggested questions to consider in assessing how well the proposal meets this criterion: How important is the proposed activity to advancing knowledge and understanding within its own field and across different fields? How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, please comment on the quality of prior work.) To what extent does the proposed activity suggest and explore creative and original concepts? How well conceived and organized is the proposed activity? Is there sufficient access to resources?"*
Broader Impacts:	*"The following are suggested questions to consider in assessing how well the proposal meets this criterion: How well does the activity advance discovery and understanding while promoting teaching, training, and learning? How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, geographic, etc.)? To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks, and partnerships? Will the results be disseminated broadly to enhance scientific and technological understanding? What may be the benefits of the proposed activity to society?"*

The current Merit Review criteria have evolved in both definition and presentation over time since their introduction in 1997, beginning with recommendations from the National Academy of Public Adminstration (NAPA) in a 2001 report comissioned by the Senate Appropriations Committee, “out of general concern to ensure accountability and responsibility in funding” (Colvard et al., 2001). Most recently, the criteria have been adjusted as a result of recommendations generated by a second Task Force in 2011 that proposed expanded and clarified guidelines for both criteria following widespread critique from and confusion of the research community at large (National Science Board, 2011). Following the recommendations of the 2011 Task Force on Merit Review, these criteria were revised for the 2013 edition of the Proposal and Award Policies and Procedures Guide (PAPPG), with the aim of addressing concerns, confusion, and inconsistencies in the application of the Merit Review criteria.

While the previous question-format criteria proposed in 1997 remained, the revised version included a stated definition of each criteria, which remains current to the available edition at the time of writing. Those definitions stated that “The Intellectual Merit criterion encompasses the potential to advance knowledge”, while the definition of Broader Impacts’ read “The Broader Impacts criterion encompasses the potential to benefit society and contribute to the achievement of specific desired societal outcomes.” (National Science Board, 2011).

Since the 2000s, there has been an increasing emphasis on Broader Impacts within the National Science Foundation. The NSF has made multiple revisions to its guidelines to clarify expectations surrounding this criterion, encouraging researchers to consider - creatively - how their work could benefit society more broadly. These revisions reflect ongoing discussions within the scientific community regarding the role of science in addressing societal challenges. A list of potential activities to meet the Broader Impacts criterion was first included in the 2013 Proposal Awards Policies and Proceedures Guide (PAPPG) (National Science Foundation, 2013), in response to repeated requests from potential applicants to clarify the NSF’s expectations around the second criterion, which and has since been expanded and added to and remains in the current edition of the PAPPG at the time of writing (National Sceince Foundation, 2024). A series of Important Notices (No. 121, 125, and 127) were particularly significant as each acted to reaffirm the importance of both Intellectual Merit and Broader Impacts to various audiences including university presidents and grant-holding organization directors, individual researchers, and future applicants (National Science Foundation, 1997, 1999, 2002).

However, balancing these criteria continues to be a contentious issue, with ongoing debates concerning the relative weight each criterion should hold in the evaluation process. During the introduction of the criteria, public comments expressed concern that excessive weighting towards Broader Impacts could

potentially undermine the strength and rigor of research proposed and funded at the NSF, with some commenters indicating they would prefer weighting Intellectual Merit 90% and Broader Impacts 10% (NSB-NSF Staff Merit Review Task Force, 1997). These concerns have endured over the years, and were echoed in both the NAPA report requested by the US Senate in 2001 (Colvard et al., 2001) and later in a survey conducted by the 2011 Task Force (National Science Board, 2011).

Today, the NSF continually refines its application and review processes to meet emerging scientific and societal needs. The constant scrutiny and periodic revisions of the merit review criteria aim to maintain the highest standards, promoting innovative and impactful scientific research. This ongoing evolution mirrors broader societal shifts in how the role of science is perceived, emphasizing not only the pursuit of knowledge but also the scientific community's responsibility to address pressing societal needs.

2. Problem Statement

In the realm of scientific research funding, the National Science Foundation (NSF) holds a strong position as one of the largest and most deeply established science funding institutions. Its grant review process employs the two core Merit Review criteria discussed earlier: Intellectual Merit and Broader Impacts. These criteria are pivotal in guiding funding decisions that affect the trajectory of scientific research and the allocation of billions of dollars. Despite their significant role in the proposal review and funding process, persistent challenges continue to surface in how these criteria are understood, valued, and implemented by various stakeholders, including researchers, Program Officers, and institutional officials.

Key stakeholders, specifically researchers applying for funding and Program Officers reviewing these applications, play pivotal roles in this process. Despite increasing support among researchers for the concept of broader impacts, as noted in

a survey conducted by the National Science Board (2011), a gap remains between their favorable attitudes and the practical challenges they face in integrating Broader Impacts effectively into their projects. Findings suggest that even well-intentioned researchers struggle with operationalizing broader impacts due to perceived imbalances in weighting favoring Intellectual Merit over Broader Impacts, challenges in measuring outcomes, and a lack of adequate support and guidance (National Science Board 2011, National Alliance for Broader Impacts, 2018). This gap highlights a critical oversight in existing literature, which often presumes that a strong belief in the value of broader impacts correlates directly with ease of implementation (National Alliance for Broader Impacts, 2018; Woodson and Boutillier, 2023).

Additionally, recent evaluations by the National Science Board's Commission on Merit Review have highlighted several critical issues: a perceived imbalance between the weighting of Intellectual Merit and Broader Impacts, the operational challenges that arise when effectively measuring and achieving impactful outcomes and ensuring accountability for research grants, and enduring concerns about the fairness and inclusivity of the review process, particularly for underrepresented researchers, including those hailing from racial and ethnic minorities, as well as women and researchers from emerging institutions. These challenges are compounded by a lack of comprehensive understanding among researchers about the criteria, and insufficient institutional support for developing impactful, integrated broader impact activities.

This dissertation investigates the complexities faced by these stakeholders in interpreting and implementing the NSF's merit review criteria throughout the research lifecycle, with a goal of providing a nuanced understanding of challenges and implementation gaps that hinder effective integration of broader impacts, even among proponents. This dissertation also aims to offer insights that could guide the

enhancement of NSF policies and support structures, ensuring that they not only foster high-quality scientific research but also promote equitable and significant societal impacts.

2.1. Key Organizational Levels

The NSF is operated by a hierarchical leadership structure with a variety of goals and focuses. The Merit Review process in particular has been stewarded by a series of task forces and committees responsible for formulating and revising the merit review criteria. These have included the 1996-1997 National Science Board (NSB)-National Science Foundation Staff Task Force on Merit Review, which proposed the Intellectual Merit and Broader Impacts criteria initially, the 2010-2011 NSB Merit Review Task Force (MR) which revised the Broader Impacts and Intellectual Merit criteria to their current form, and the current 2023-2024 NSF-NSB Commission on Merit Review (MRX) which is ongoing as of Fall 2024. These top-level entities play a crucial role in shaping the overarching policies and strategic directions of the NSF through a combination of strategic oversight, expert input, and public comment. The decisions of NSF officials broadly, and the various Merit Review Task Forces over time specifically, have been informed by a wide array of considerations, including scientific priorities, societal needs, governmental oversight, and other regulatory requirements. As a result, the policies they establish serve as the foundation upon which the entire Merit Review process is built and must be considered accordingly in any analysis of the topic.

2.2 Program Officers

Shifting to an individual level, Program Officers occupy a pivotal position within the NSF, and have traditionally functioned as the linchpin between the top-level decision-makers and the researchers submitting proposals. These individuals,

often researchers in their own right who have rotated into a Program Officer role for a set period of 1-2 years, are tasked with interpreting and operationalizing the policies set forth by the NSF leadership during the review process. In doing so, they engage in a delicate balancing act, striving to meet the strategic goals of the organization while also addressing the practical concerns and needs of the research community.

Mazuzan (1992) notes that reviewers being strictly advisory to the program directors and later Program Officers maintained the principle that decisions regarding funding supplied by federal budgets appropriated to the NSF must ultimately be controlled by federal officials. A significant part of the program directors' roles, also according to Mazuzan, was to not only be deeply involved in the field in question, but also to identify reviewers who know the field best. The role of Program Officers as apparent decision makers and arbiters around grant applications and criteria stands as another key side of the system to be understood. The apparent complexity and 'black-boxing' of the the NSF criteria, and the role of Program Officers in that process, is potentially a result not of intentional obscuration but rather of systemic complexity, jargon, and nuance that many researchers may not feel prepared to navigate, as evidenced by the variety of tools, workshops, and explainer documents that exist to guide individuals through the process (National Alliance for Broader Impacts 2018).

Program officers must navigate a complex landscape, providing guidance to researchers, coordinating review panels, and ensuring that proposals are evaluated fairly and consistently. In contrast to individual researchers who are being assessed on a single proposal at a time, NSF Program Officers engage with numerous proposals in a year and work continually with Principal Investigators (PIs) over the lifetime of the funded projects. Their role is critical in ensuring that the merit review process is both rigorous and adaptable to the diverse needs of the scientific

community, whilst also maintaining a longer-term 'birds-eye view' of their research areas that stretches beyond a specific project or moment in time. Yet, as most Program Officers come from the wider research community, their experience and approaches offer a valuable means of linking experiences and expectations between the institutional and individual actors within the funding system.

2.3 Researchers

At the foundational level are the individual researchers and scholars who develop and submit proposals to the NSF. These scientists are on the front lines of the merit review process, tasked with interpreting the criteria and demonstrating how their work meets both the intellectual merit and broader impacts requirements in a way that is both intelligible and persuasive to review panels and Program Officers alike. This translation from theory to active practices leads researchers to stress-test the criteria, guidelines, and any external support resources offered in a vast breadth of ways.

Particularly given the breadth of allowable activities afforded by the NSF in their definition of Broader Impacts, and the need to craft statements that can 'sell' a proposal to a review panel, individual scholarly understandings of Broader Impacts may not necessarily reflect the entire scope of activities and research elements that actually impact communities, publics, and/or colleagues (Frodeman et al. 2013). Instead, these individual researcher's definitions of 'Broader Impacts' may reflect particular narratives of the role of science in society, the research process, and their identity as a researcher that exert significant influence over the resulting outputs. As such, understanding these narratives around how researchers and scholars define 'Broader Impacts' both in theory and in their own practice will help accurately contextualize the resulting statements and the breadth of both proposed and actually executed Broader Impact activities.

3. Research Questions and Summary of Studies

The three papers involved in this dissertation will explore how individual researchers, Program Officers, and institutional officials perceive and respond to the NSF's criteria, shedding light on the challenges they face in aligning their scientific goals with the expectations of the funding agency. NSF projects span an enormous range of disciplines, topics, and possible modes of impact, with an enormous number of project applications funded - and even more submitted - every year. Regardless of researchers' – and NSF Officials' – own personal perceptions of the Broader Impacts criteria, it remains a requirement for attaining funding at one of the most well-known scientific funding agencies both in the US and globally; the decision to elevate Broader Impacts to a requirement will necessarily have knock-on effects on the research community at a national scale.

As such, this dissertation aims to consider the more specific space of the current Merit Review criteria, asking the question: **How and why do stakeholders engaging with the NSF - either as researchers, Program Officers, or institutional officials - construct and value research impact, particularly as a function of the NSF's Broader Impact criteria?**

A stated goal of the Broader Impact criteria is to increase the rate at which scientific research in the United states provides societal benefits, particularly in terms of science education and outreach (National Science Foundation, 2024; Holbrook, 2012). The implementation of the Broader Impacts criterion asks scientists and researchers to move beyond creating just novel research outputs, and consider which audiences receive and can benefit from that knowledge. Moreover, the format of this criteria asks how, why, and when they themselves as the principal researchers will support and further that impact (Frodeman et al. 2013, Roberts 2009).

Table 2. Research Questions and research methodology applied.

	Research Question	**Methodology**
Main Question:	How and why do stakeholders engaging with the NSF - either as researchers, Program Officers, or institutional officials - construct and value research impact, particularly as a function of the NSF's Broader Impact criteria?	Results from both sub-questions were compiled and considered across both individual and institutional perspectives. A mixed method approach pairing quantitative analysis with qualitative coding for all data types was applied.
Research Question 1 (RQ1):	How and why do individual researchers and scholars understand and value 'Broader Impacts' as an aspect of their own research and experiences?	Survey of scholarly network affiliated with R1 university; questions focused on familiarity with and understanding of Merit Review and recounting previous activities and experiences. Semi-structured interviews with researchers who have submitted NSF grants, selecting for a variety of disciplinary backgrounds, career stages, and representation. A textual elicitation portion was incorporated presenting the NSF criteria during each interview.
Research Question 2 (RQ2):	How do NSF Program Officers and officials structure and facilitate impact through their role within the NSF application review process?	Analysis of NSF-NSB Task Force on Merit Review open meetings using LLM-Assisted transcription, with additional analysis of official documents as needed.

To this end, the overarching research question can be further broken down into two research questions, as seen in Table 2, beginning with: **How and why do individual researchers and scholars understand and value 'Broader Impacts' as an aspect of their own research and experiences?** This first research question (RQ1) is significant because any disconnect between the institutional perspective, either at the level of the NSF or specific universities, and the actual perspective of individual scholars within that organization as to what, why, and how Broader Impacts relate to research efforts may undermine their practice.

Having established a foundational understanding of Broader Impacts in theory and in practice for individual researchers, it is also critical to understand the

ways in which value is structured in the current NSF proposal review process. As such, this dissertation also asks a second research question (RQ2): **How do NSF Program Officers and officials structure and facilitate impact through their role within the NSF application review process?** RQ2 was approached via the timely avenue of the (ongoing at the time of writing) 2023-2024 NSB-NSF Task Force on Merit Review.

In this dissertation, RQ1 and RQ2 are both addressed using a mixture of survey results and semi-structured interviews, as noted in Table 2. The 900+ researchers contacted were all participants in a scholarly network at a major public R1 university; participation in this scholarly network was chosen as a criteria for inclusion largely for ease of access. However, the nature of the scholarly network, focused on global and local challenges, may overrepresent researchers and scholars who self-identify as generally supportive of Broader Impacts. Researchers who completed the survey were also offered the opportunity to participate in follow-up semi-structured interviews to discuss their understandings of and perspectives on the Merit Review criteria generally and Broader Impacts specifically in more detail. Former NSF Program Officers were identified and contacted through a mix of personal networks and introductions via email; interview questions for both participant groups were similar but adjusted to reflect their particular roles as individual researchers or former NSF Program Officers. Both long-form survey responses and interview responses were analyzed using qualitative coding.

While analysis of former Program Officers' perspectives was initially combined with those of institutional officials as part of understanding the institutional perspective on the Merit Review criteria in Research Question 2, initial analysis of the data indicated that the former Program Officers' perspectives were far more individualistic and as such more closely aligned with the perspectives of individual researchers. As such, the final analysis in this dissertation is restructured

to approach both researcher and Program Officer perspectives and understandings of Broader Impacts within the space of RQ1, as can be seen in Table 3.

Table 3. Guide to papers in this dissertation.

	Title	**Topics**	**Research Question Addressed**
Chapter 2 (Paper 1)	The Weight of Merit: Balancing Policy and Practice in the NSF's Merit Review Criteria	Per-project 'weighting' of the Merit Review criteria by individual researchers; role of grants and awards in tenure and promotion	RQ1
Chapter 3 (Paper 2)	Broadening the Scope of Broader Impacts: Evolving Guidelines, Partial Success and Ongoing Barriers	Types of Broader Impacts activities; Broader Impact Centers	RQ1
Chapter 4 (Paper 3)	Broader Impacts, Officially: The Proceedings of the 2023-2024 NSB-NSF Task Force On Merit Review	'Weighting' of Broader Impacts during portfolio balance; Broader Impacts implementation and accountability	RQ2

The historical imbalance between criteria weighting when reviewing and accepting proposals, specifically how the Broader Impacts criteria are represented and weighed against Intellectual Merit, and how institutional and individual funding considerations drive decisions as to what kinds of Broader Impacts activities are realistic and achievable has a high influence on the landscape of Broader Impacts. Moreover, all three stakeholder groups in this research - individual researchers, Program Officers, and the institution of the NSF at large - have varying degrees of commitment to and control over each of these factors.

Given this, following both the main research question and RQ1 and RQ2 (Table 2), each of the three papers in this dissertation was developed along an individual-institutional perspective axis, as seen in Figure 1.

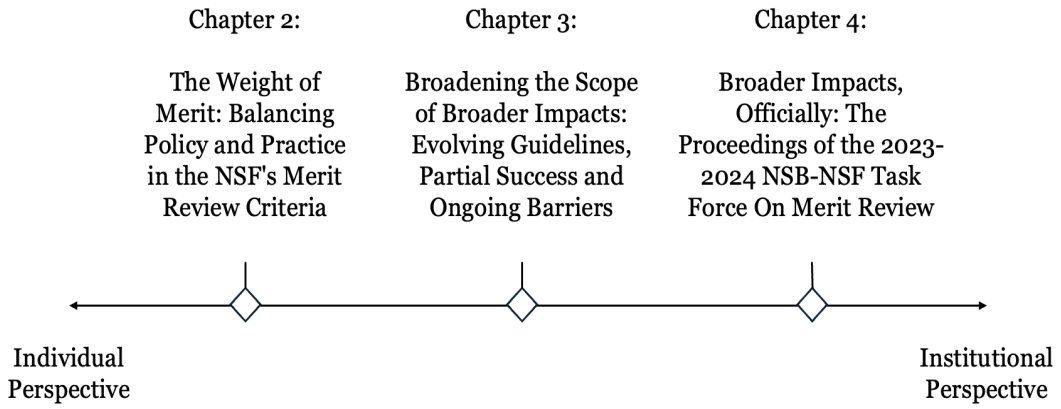

Figure 1: Position of each paper in this dissertation along an axis of Individual-Institutional perspective.

The first paper of this three-paper dissertation, in Chapter 2, is primarily focused on enduring concerns around the weighting of Broader Impacts versus Intellectual Merit in the proposal review and award process. Using researcher survey data, and interviews with both individual researchers and Program Officers (Table 3), Paper 1 considers how individual researchers understand Broader Impacts in their own work - both as a personal value and aspect of their research process, but also as a factor in not only the grant review and approval process but their own individual career trajectory through the promotion and tenure process.

Results of this paper indicate that weighting remains an enduring concern among both researchers and Program Officers that has been maintained from the introduction of the criteria in 1997 through to the current day. Furthermore, survey and interview data results indicate that even while attitudes towards Broader Impacts may be generally improving, Broader Impacts are still seen as fundamentally less important than Intellectual Merit, both among individual researchers themselves, as well as within their research institutions and centers, directly shaping a reduced role for Broader Impacts work in the promotion and tenure process.

The second paper, in Chapter 3, brings together the individual and institutional perspectives among researchers and Program Officers with regard to types of Broader Impact proposed and executed through a mix of survey and interview data (Table 3). Survey responses concerning types of Broader Impacts activities researchers had performed indicated every type was represented. Researchers generally focused on their own broader impacts activities at an individual level.

In contrast, this paper proposes that former Program Officers introduce institutionally-informed insights, offering higher-level, 'birds-eye-view' perspectives on the range and depth of broader impacts they observed in proposals. In particular, interview responses from former Program Officers indicate some degree of frustration with what they saw as a prevalent trend among researchers to rely heavily on broadening participation and educating undergraduate and graduate students as 'default' broader impacts activities.

Paper 2 also bridges the individual and institutional perspectives through the lens of Broader Impacts Research Centers, which typically serve to connect individual researchers with a larger scope of Broader Impacts activities, opportunities, and research projects within their own institutions. As such, researchers' and Program Officers' discussions of the role of Broader Impacts Centers, both positive and negative, offer insights into the ways they begin to approach positioning their individual efforts within a larger institutional network of activities and impact.

Lastly, the third paper, in Chapter 4, focuses on institutional perspectives on Broader Impacts through the lens of the National Science Foundation. Here, the National Science Foundation as an institution is accessed though the open meeting proceedings of the NSB-NSF 2023-2024 Task Force on Merit Review. In considering transcripts of the open meeting proceedings of this Task Force, Paper 3 offers an unique glimpse into the interior workings of the National Science Foundation and the

ways institutional officials and representatives understand, construct, and reinterpret the Merit Review criteria in ways that become the official position of the National Science Foundation as an institution. The results of Paper 3 place questions of weighting, implementation, and accountability in the context of unique expectations and tensions present only at the larger institutional scale, and highlight how such concerns are both maintained and altered when shifting from individual to institutional perspectives.

Following all three papers, the conclusion to this dissertation discusses how stakeholders' concerns, expectations, and varying understandings of the Merit Review process generally, and the Broader Impacts criterion specifically, come together and diverge. Additionally, with potential changes to the Merit Review process expected at the conclusion of the NSB-NSF 2023-2024 Task Force on Merit Review proceedings, select possible future changes and considerations are also discussed. As this dissertation is being written in 2024, with changes anticipated to the Merit Review process following the release of recommendations from the 2023-2024 NSB-NSF Task Force on Merit Review later this year, an understanding of the current Merit Review framework is crucial for researchers and institutions to effectively anticipate and navigate future developments in the process.

Through this process, this exploratory dissertation seeks to uncover how the language and policies established by the NSF's top echelons are translated into practical guidelines and actions at the ground level by individual researchers, shaping the process of scientific research and scholarship from the top down, and the bottom up simultaneously. By focusing on scaling from personal narratives and experiences up to institutional perspectives and expectations, this dissertation aims to present a nuanced view of the merit review process, highlighting the complexities and challenges involved in balancing Intellectual Merit with Broader Impacts, and with creating and maintaining a robust cycle of research funding, output, and impact.

4. References:

Colvard, James, Bostrom, Carl, England, Mary Jane, & Hale, Sandra. (2001). *A Study of the National Science Foundation's Criteria for Project Selection*. National Academy of Public Administration. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=dbeb6cd123afda1f77369490fd27ffcb3e2df6d9

Chubin, D. 2007. http://www.issues.org/24.1/forum.html.

Davis, M., & Laas, K. (2013, October 24). "Broader Impacts" or "Responsible Research and Innovation"? A Comparison of Two Criteria for Funding Research in Science and Engineering. SpringerLink. https://doi.org/10.1007/s11948-013-9480-1

European Research Council. (2023). *ERC at a glance*. ERC. https://erc.europa.eu/about-erc/erc-glance

Holbrook, J. B. (2012). Re-assessing the science–society relation: The case of the US National Science Foundation's broader impacts merit review criterion (1997–2011). *Peer Review, Research Integrity, and the Governance of Science–Practice, Theory, and Current Discussion*, 328-362.

Mazuzan, G. T. (1992). " Good Science Gets Funded..." The Historical Evolution of Grant Making at the National Science Foundation. Knowledge, 14(1), 63-90.

National Alliance for Broader Impacts (NABI). (2018) The current state of broader impacts: advancing science and benefiting society. Columbia (MO): National Alliance for Broader Impacts. https://www.broaderimpacts.net

National Science Foundation (1997). *Important Notice No. 121* (in121). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in121.txt

National Science Foundation (1999). *Important Notice No. 125* (in125). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in125.pdf

National Science Foundation (2002). *Important Notice No. 127* (in127). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in127.pdf

National Science Board. (2011). National Science Foundation's Merit Review Criteria: Review and Recommendations (NSB/MR-11-22). National Science Board. https://www.nsf.gov/pubs/issuances/in121.txt

National Science Foundation (2013). Proposal and Award Policies and Procedures Guide (NSF 13-1, NSF13001). Washington, DC. Accessed 13 June 2024 at https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/nsf13_1.pdf

National Science Foundation (2023). *Proposal and Award Policies and Procedures Guide* (NSF 24-1, NSF24001). Washington, DC. Accessed 22 June 2024 at https://nsf-gov-resources.nsf.gov/files/nsf24_1.pdf

National Science Foundation. (2024). *NSF Budget Requests to Congress and Annual Appropriations*. https://new.nsf.gov/about/budget/fy2021

National Science Foundation. (n.d.). *Merit Review: Facts*. Retrieved September 21, 2022, from https://www.nsf.gov/bfa/dias/policy/merit_review/facts.jsp

Rothenberg, M. (2010). Making judgments about grant proposals: A brief history of the merit review criteria at the National Science Foundation. *Technology & Innovation*, *12*(3), 189-195.

Woodson, T., & Boutilier, S. (2023). From intent to impact—The decline of broader impacts throughout an NSF project life cycle. *Research Evaluation*, *32*(2), 348-355.

CHAPTER 2

THE WEIGHT OF MERIT

The Weight of Merit: Balancing Policy and Practice in the NSF's Merit Review Criteria

1. Introduction

> "Intellectual Merit: The Intellectual Merit criterion encompasses the potential to advance knowledge."

and

> "Broader Impacts: The Broader Impacts criterion encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes." (NSF, 2021, III.A.2)

Intellectual Merit and Broader Impacts were introduced as the two defining Merit Review Criteria at the National Science Foundation in 1997, following a long proposal and public comment process (Rothenburg, 2010). At the time of their proposal, both criteria, and particularly the Broader Impacts criterion, were subject to significant scrutiny and critique. Reactions from the required public comment period at that time ranged from concerns about ambiguity around the Broader Impacts criterion and the topic of 'diversity', to questioning how the NSF expected to convince the scientific community to support and follow the new criteria; the Task Force on Merit Review assigned to carry out and report on the process provided a series of recommendations to address each issue. However, by far the largest issue raised in 1997 was what the Task Force termed the "Weighting issue" - that the majority of public commenters found Intellectual Merit to be "much more important" than Broader Impacts, and expected the criteria to be reviewed and weighed accordingly (NSB-NSF Staff Merit Review Task Force, 1997).

As the current Merit Review criteria approach their 30th anniversary, understanding the long-term perceptions of Intellectual Merit and Broader Impacts

by the scientific community is ever more important. However, little research has directly addressed these questions at the individual level since the Merit Review criteria were last reassessed in 2012 (Frodeman et al., 2013), or considered how those perspectives and concerns may have evolved as both criteria age. To that end, this study seeks to investigate the roots and ramifications of this imbalance, providing insights that can guide future policy at the NSF and potentially pave the way for cultural shifts within academic institutions to better align with the dual objectives of advancing knowledge and benefiting society.

As such, this paper considers the perspectives of two key stakeholder groups in 2024: researchers submitting NSF grants, and former NSF Program Officers who were responsible for guiding proposals through the peer-review process under the current criteria. Survey data and semi-structured interviews were analyzed to identify key themes and trends in both stakeholder groups' understandings of the Merit Review criteria, finding that while the majority of respondents were strongly supportive of Broader Impacts, weighting continues to be a significant concern for researchers and Program Officers alike, with downstream influences on researchers' career advancement.

Furthermore, tensions emerged between participants' internal valuation of Broader Impacts and a more widely acknowledged value of Intellectual Merit as 'good science', echoing the concerns noted in 1997. Lastly, this paper concludes by considering the challenges facing both individual researchers, and the NSF more broadly as it looks to assess the Merit Review in the coming years. By exploring the experiences and perspectives of both researchers and Program Officers, this research aims to illuminate the underlying causes of the perceived imbalance and start a path towards achieving a more equitable evaluation process that values both Intellectual Merit and Broader Impacts.

2. Background and Literature Review

2.1 Historical Changes

The NSF has always maintained some mandate towards societal benefit, though not always at the forefront of its institutional culture. The National Science Foundation was created and endowed in 1950, following four research criteria: researcher performance competence, the intrinsic merit of the research, the utility or relevance of the research, and its effect on the infrastructure of science and engineering (Rothenberg, 2010). Starting in the 1980s, societal benefit and research impact began to be more tightly integrated into the review and award process (McCullough, 1989), though not without concern, culminating with its codification into the current Merit Review Criteria in 1997. At that time, after an extensive public comment period and shepherded by a Task Force dedicated solely to the issue, those four criteria were streamlined to the current two: Intellectual Merit, and Broader Impacts (NSB-NSF Staff Merit Review Task Force, 1997).

However, uptake was neither swift nor cohesive; in 2002, the NSF issued a Dear Colleague letter informing potential applicants that any application lacking a Broader Impacts statement would be rejected (National Science Foundation, 2002). This letter hinted at a, at the time unspoken but broadly recognized, reality that the NSF had been facing a significant lack of uptake regarding the Broader Impacts criteria (Kamenentsky, 2012; Frodeman et al., 2013). Later, the NSF further refined documentation and guidance around the Broader Impacts criterion in 2012 in order to clarify expectations and provide greater guidance to applicants (National Science Foundation, 2013).

Today, that guidance primarily takes the form of the Proposal and Award Policies and Procedures Guide (PAPPG), which is the main official NSF document detailing how the Merit Review criteria are to be understood and implemented. Concerning the Merit Review criteria, the guide states:

> "The two merit review criteria are listed below. Both criteria are to be given full consideration during the review and decision-making processes; each criterion is necessary but neither, by itself, is sufficient. Therefore, proposers must fully address both criteria." (NSF, 2021, III.A.2)

The relative emphasis of each criterion against the other, described here as 'full consideration', has been a long standing point of concern, dating all the way back to the introduction of the revised criterion in 1997. The final report by NSB-NSF joint task force in charge of the revision found that 'weighting' was the top 'central issue' raised during the public comment period, noting:

> "Many respondents expressed concern that adopting the new criteria will lead to a decline in NSF's standards of excellence; i.e. "excellent research with ok relevance" will be equated with "ok research with excellent relevance." Others stated that, for research proposals, Criterion #1 is much more important than Criterion #2, and should be weighted accordingly (some suggested 90/10)." (NSB-NSF Staff Merit Review Task Force, 1997).

This historical apprehension regarding the balance between Intellectual Merit and Broader Impacts underlines the uncertainties and debates that resonated at that time. Follow-up NSF guidelines and Dear Colleague letters addressed these concerns in a variety of ways, including providing increased guidance (NSF, 1997; 2005; 2012) and annual reviews of the implementation of the Merit Review criteria, a practice that has been in place since 1977 (NSF 2005). Yet today, echoes of those concerns continue to persist among researchers and faculty with respect to the roles of and expectations around the Merit Review criteria.

2.2 Current Research and Critiques

The "Weighting issue" raised in 1997 continues to be a major concern amongst researchers and grant applicants in the current research landscape. Multiple significant discrepancies exist between the intended "equal consideration" (NSB-NSF Staff Merit Review Task Force, 1997; National Science Foundation, 2024) of these criteria and their practical implementation, an uneven relationship that has been

noted both anecdotally and in research (Frodeman et al., 2013; Woodson and Boutilier, 2023).

The Broader Impacts criterion was not without critique, as already discussed, and in fact during the initial comment period around the introduction of the Intellectual Merit and Broader Impacts criteria, as well as the later 2012 reassessment, much of the public comment submitted was concern or critique (NSB-NSF Staff Merit Review Task Force, 1997; NSB Task Force on Merit Review, 2011). Researchers and faculty have spoken out during the required public comment periods required by each Task Force, with concerns ranging from the clarity of the criteria and the mechanisms by which they enforce societal benefits, to the very nature of the role of the NSF and science as a lever for societal benefit.

Since their introduction and through subsequent reassessments, a significant amount of public feedback has pointed to the vagueness and ambiguity of the Broader Impacts criterion. Holbrook and Frodeman (2012) have documented persistent resistance within the scientific community, attributing it to the criteria's perceived irrelevance to the research process, as well as unclear expectations set by the NSF, which many researchers view as more of a burden than an aid to fostering genuine societal engagement. Lok (2010) and Holbrook and Frodeman (2012) both criticized the NSF for the vagueness and expansivity of the Broader Impacts criterion, arguing that the NSF's efforts to allow flexibility in Broader Impacts activities are often perceived by researchers as a lack of clarity, adding to the burden rather than fostering meaningful engagement with broader societal impacts. Lok (2010), in particular, argues that the lack of specificity frustrates applicants who themselves would prefer clearer, more explicit guidelines for proposing Broader Impacts activities.

Critiques have extended beyond the criteria's clarity to the fundamental approach of the NSF in leveraging societal benefits through the peer review process.

Sarewitz (2011) argues that the Broader Impacts criteria are politically motivated, specifically pointing at the nine listed national goals as arbitrarily defined and proposes a selection of other topics that could also merit national focus. He challenges the efficacy of using individual research projects as a means to align NSF research with broader national goals, argues that such an approach will often lead to exaggerated claims about the societal impact of research, rather than substantive contributions, and instead suggesting instead that a higher-level, strategic programmatic approach would be more effective. Echoing this sentiment, Barry Bozeman (2009) highlights historical resistance to the Broader Impacts criterion and argues that peer review is an inappropriate mechanism for mandating societal benefits, instead advocating for a shift towards a policy-focused, institutional-level approach that could more effectively harness the social impacts of scientific endeavors.

Together, these critiques underscore a broader concern: the Broader Impacts criterion, as currently structured and implemented may serve to complicate the research process with unclear guidelines and inappropriate evaluative mechanisms, but also may fail to achieve their intended societal impacts. This ongoing discourse, ebbing and surging over the last thirty or more years, suggests a need for a fundamental reevaluation of how the NSF's goals for societal engagement are articulated and enforced.

2.3 Criteria In Practice

Applying the Merit Review criteria is often much more nuanced in practice than on paper. Woodson and Boutilier (2023) found that more Broader Impacts activities are proposed than are actually conducted, and largely drop off over the course of a research project's life cycle. Roberts (2009) found that proposals that did directly reference societal benefits rarely did so in conjunction with concrete

dissemination plans. It is unclear whether that discrepancy was due to considering the act of conducting research itself as an intrinsic societal benefit (thereby assuming that further explanation or dissemination is unnecessary), or a lack of experience articulating broader dissemination plans. Societal benefit absent dissemination plans may also reflect a tendency to include generic statements of potential impacts more as a formality or theoretical ideal rather than as a practical component of the research strategy or actual project execution (Salet 2018).

Particularly following the clarification of the Broader Impacts criterion for fiscal year 2012 (NSB Task Force on Merit Review, 2011), numerous studies have considered the types of activities proposed in NSF grant proposals (Wiley 2014, Nagy 2016) often finding that most proposed activities emphasized teaching and training activities, including conference presentations, training undergraduate students, and other Broader Impacts activities focused on peer-to-peer engagement within academia (Wiley, 2014; Kamenetzky, 2013; Roberts, 2009). Next most common are Broader Impacts activities focused on broadening dissemination to non-academic K-12 audiences (Wiley, 2014; Nadkarni and Stasch, 2013); these trends tend to hold true both within specific NSF divisions as well as across divisions (Nadkarni and Stasch, 2013; Watts et al., 2015).

The types of Broader Impacts activities proposed by researchers and the emphasis on 'typologies' of Broader Impacts stands in contrast to the hidden curriculum around Intellectual Merit - little to no research focuses specifically on the other half of the Merit Review criteria, or explores how intellectual merit is understood by researchers. The 2011 NSB Task Force on Merit Review found that while Intellectual Merit was "well-understood by the community and NSF staff" (pg. 8), what research exists largely positions Intellectual Merit as a foil to understanding or critiquing Broader Impacts.

2.4 Tenure and Promotion

The literature on tenure and promotion processes and criteria within academia are primarily focused on analyses of efficiency and interdepartmental dialog (Schimanski and Alperin, 2018; Ambroze and Cropanzano, 2008), and concerns around equality and bias (Perna, 2001; Oleschuk, 2020), particularly with regard to women (Winkler, 2000) and underrepresented minority faculty (Perna, 2001). Scholarship and publication rates also receive significant attention in this space, with publication pressure and the informal phrase 'publish-or-perish' commonly used to describe a culture within academia that rewards a high rate of publications and research work (Van Dalen and Henkens, 2012), while leading to anxiety, stress, and potentially undermining research quality (Tijdink, Verbeke and Smulders, 2014) and increasing researcher bias (Fanelli, 2010).

However, another aspect of the tenure and promotion process that directly connects to Broader Impacts is the role of public engagement, outreach, and 'non-scholarly' work in tenure and promotion considerations. These activities, which may variously be encompassed in the umbrella of Broader Impacts activities, are not always recognized or rewarded (O'meara, 2010), and the language used to describe such efforts - from engaged research, to research impact, and more - and the accompanying reward scales, are highly varied across institutions (Doberneck, 2022).

Some research has emerged calling on universities and organizations to recognize and reward researcher's impact-focused work during annual evaluations and tenure proceedings (Bouwma-Gearhart et al., 2021). However, with regard to Broader Impacts specifically, no research was identified that considered the National Science Foundation's criterion in the tenure and promotion process specifically. However, given general findings that traditional publication and research metrics outweigh public dimensions in career decisions (Alperin et al., 2019; Bouwma-Gearhart et al., 2021; O'meara, 2010), it is reasonable to conclude that Broader

Impacts works as a specific area of research impact and public engagement work is subject to similar pressures.

3. Methodology:

3.1 Survey

A mixed methods survey was conducted in order to establish a preliminary understanding of researchers' familiarity with, perceptions of, and experience with the NSF merit review criteria. Researchers at a large public R1 institution were identified via membership in a highly interdisciplinary scholarly network focused on addressing global concerns and challenges. The alignment of network members with pro-Broader Impacts perspectives is acknowledged; however, given the scant literature addressing the perspectives and questions of those who are themselves supportive of Broader Impacts, the survey population was deemed valuable for this study. Members of this network were contacted via the official network mailing list for survey participation. In total, 922 researchers were contacted, with 63 completeing the survey and 11 completing an interview, as shown in Figure 2.

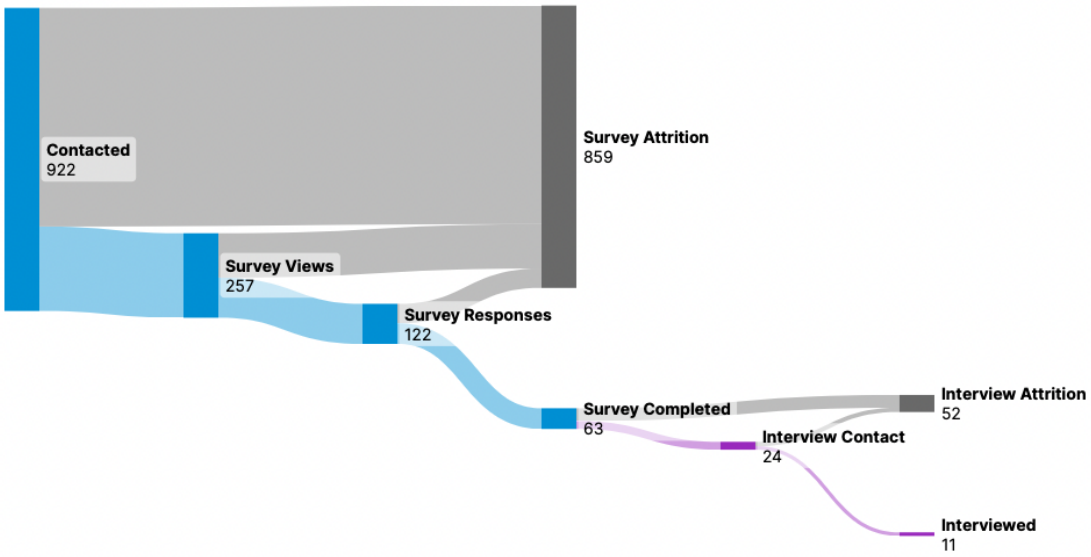

Figure 2. Researcher recruitment and attrition during survey and interview phases.

The survey was conducted digitally using the QuestionPro survey platform, and included 25 questions, with 14 short-form Likert scale questions and 11 open-ended long-form questions (See Appendix A). The intended duration was 30 minutes or less; duration and survey questions were tested prior to dissemination with two pilot testers from within the survey population; the survey protocol was approved by the Arizona State University Institutional Review Board (Appendix E). The survey was conducted anonymously, but concluded by offering participants the opportunity to share their contact information if they wished to participate in a longer follow-up interview.

3.2 Semi-Structured Interviews

Semi-structured interviews were conducted with both researchers and former Program Officers. Researchers who shared their contact information in the survey were recruited via email. Separately, current and former Program Officers were contacted for research interviews using personal and professional networks. While both current and former Program Officers were contacted to participate, only former Program Officers opted to be interviewed; as such their participant group may be referred to as Program Officers or former Program Officers interchangeably. All interviews with both researchers and Program Officers were scheduled for one hour and conducted digitally via videoconference. Interview protocols for both groups followed a similar order and length, with questions modified only to refer to the participant's role – Researcher or Program Officer – and can be found in Appendices B and C. Both interview protocols were approved by the Arizona State University Institutional Review Board (Appendix E), and all interviews were conducted between January and March 2024.

3.3 Artifact elicitation

For the purpose of this research, the artifact elicitation method was implemented during both the researcher and Program Officer interviews (Douglas et al., 2015). Two virtual text artifacts were designed using Microsoft PowerPoint as slides that respectively defined the Intellectual Merit (Figure 3) and Broader Impacts criteria (Figure 4) using the official definitions provided in the current edition of the PAPPG (National Science Foundation, 2024).

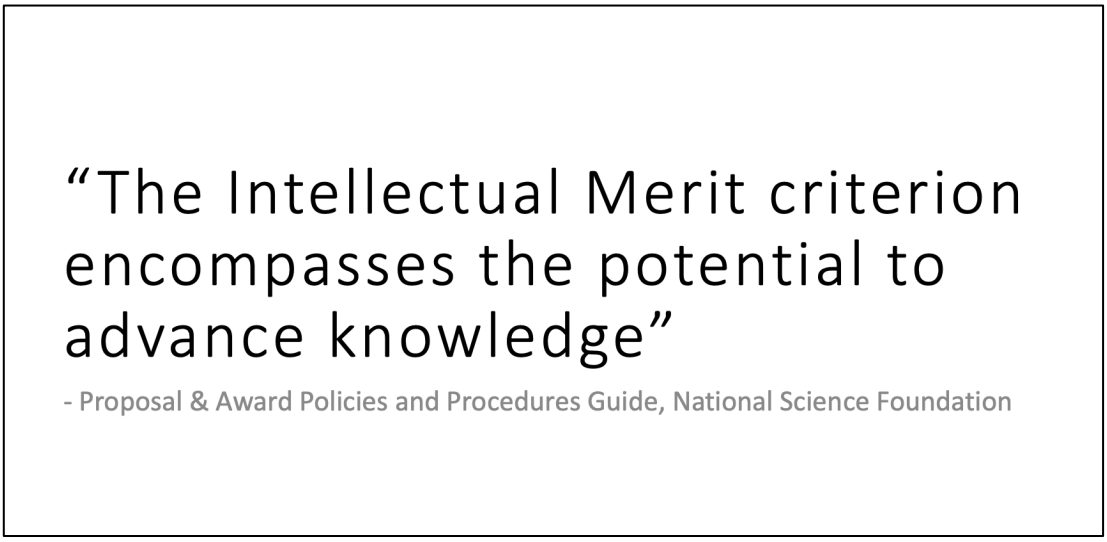

Figure 3. Intellectual Merit criterion artifact shown to all participants.

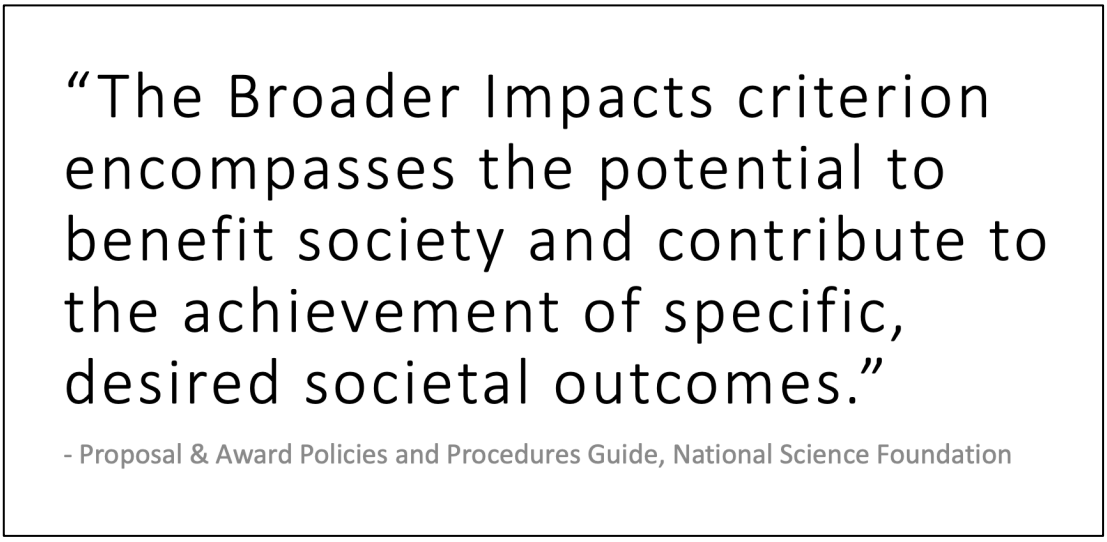

Figure 4. Broader Impact criterion artifact shown to all participants

Participants were first shown the slide on Intellectual Merit in Figure 1 during the artifact elicitation portion of the interview (Appendix B) and asked to describe their understanding of the criteria. Participants were also encouraged to describe what it meant to them and reflect on its relevance to their own research or professional practice, and discuss any experiences related to these criteria. Lastly, participants were asked to reflect on if they believed their understanding of the criteria was shared with colleagues in their field or aligned closely with the NSF's understanding and expectations. This process was then repeated with a second slide showing the Broader Impacts criterion (Figure 4) and the same questions were asked.

These slides served as focal points during the interviews, prompting participants to discuss and reflect upon each criterion in detail. By presenting these definitions in a structured, visual format, we aimed to provide a reliable touchpoint for interviewees to refer to during their answers, and to standardize the information provided to each participant while still allowing for individual interpretations and responses to be freely expressed. This sequential presentation and discussion helped in tracing how participants differentiated and related the two criteria.

The use of artifact elicitation in this context not only enriched the data collected by anchoring the discussions in the specific concepts in question, but also enabled a focused exploration of how these foundational criteria are conceptualized and operationalized by researchers and Program Officers. This methodological approach proved instrumental in revealing the depth of the participants' engagement with, and the personal and professional significance they attribute to, these key assessment criteria.

Audio recordings were saved to a secure cloud storage location and transcribed for analysis. As the primary focus of this work was not the mechanics of speech but rather the substantial content of the discussion, particularly interviewees'

perceptions of the Merit Review criteria, a denaturalized transcription approach was taken (Oliver et al., 2005). All transcriptions were then uploaded to Dedoose for qualitative coding and analysis.

3.4 Flexible qualitative coding

Deterding and Waters' "Flexible Coding of In-depth Interviews: A Twenty-first-century Approach" (2018) was used for interview analysis, the stages of which are shown in Figure 5. This methodology was chosen for this project for several key reasons. First, existing literature indicates that scholarly and institutional perspectives on the NSF merit review criteria involve nuanced, multi-dimensional aspects, both practical and emotional, that could not entirely be anticipated during the initial interview protocol development. As flexible coding allows for modifications in the coding scheme as more understanding is gained about the data, this would assist in ensuring that the analysis remained relevant and comprehensive.

Additionally, given the expertise of the participant population, who are more accustomed to being on the other side of the research process, these interviews and the accompanying coding process exemplifies the challenges of 'studying up' (Gusterson, 1997). The flexible coding methodology was therefore chosen with the goal of explicitly documenting these dynamics, thereby facilitating a comprehensive accounting of them within the analysis and subsequent discussions, acknowledging the hierarchical academic structures influencing the data interpretation process. An overview of the stages of the flexible coding process, as applied to this research, is shown in Figure 5.

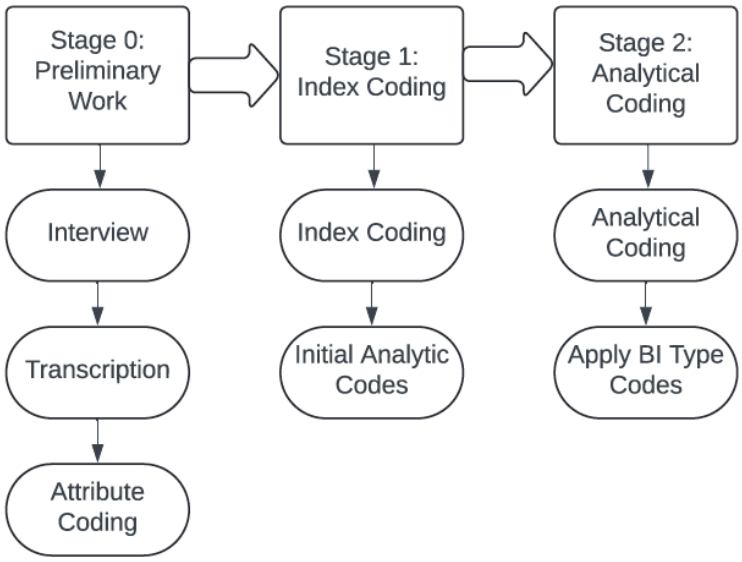

Figure 5. Stages of Deterding and Waters (2018) flexible coding approach, adapted to this study.

Following the stages in Figure 5, attributes, index codes, and analytical codes were applied to the data. Attributes were based on participants' characteristics, such as their self-identified primary research field. All participants were assigned anonymized IDs (R## for researchers, P## for Program Officers) used throughout the analysis and referencing (see Table 1). Index codes, derived from research questions and designed to provide a broad overview of the data (Deterding and Waters, 2019), were then applied. A participant's response to a specific question was coded as one unit (see Table 4).Lastly, in the final stage of analysis, analytical codes were applied to participants' responses; these codes were primarily in vivo codes that worked from participants actual phrasing to develop analytical code categories. Attributes and select index and analytical codes used can be found in Table 4.

Table 4. Example Index and Analytical codes as applied.

Code Type	Application or Definition	Coding Example
Index Codes		
Understandings of BI	Interview Question, asked during Artifact Elicitation with Broader Impacts criterion definition displayed: "Personally, how do you understand Broader Impacts in your own research?"	"But broader impacts for me is a range of different societal benefits, whether it is education of the general populace, engagement with the general populace helping people understand what science is and that what is truth?" (R03)
Understandings of IM	Interview Question, asked during Artifact Elicitation with Intellectual Merit criterion definition displayed: Personally, how do you understand intellectual merit in your own research?	"The intellectual merit basically is about the research. It's about what is the research question? And based on that research, what will the answer get us? I always tell people the simple question is who cares? Does anybody care that you're going to do this research? Because if nobody cares, then don't do it, because what's the point? So that's the way I tend to look at intellectual merit." (P06)
Support	Interview Question: How have you been or how could you be further supported in proposing or doing Broader Impacts?	"However. I will say that I in the grants that I've submitted, I do feel like I'm doing this alone, like I'm... I wish there was a little bit more hand-holding, let's say, for the beginning, because this isn't- I don't always feel I have any support from any of my supervisors." (R09)
Analytical Codes		
Weighting	How are BI and IM weighted (in terms of importance, value, necessity, function, and feasibility) in relation to or against each other? Includes statements around how the NSF is perceived to weigh each criterion, or how the participant or their colleagues might weight the criteria.	"And of the two criteria, the Intellectual Merit and Broader Impacts that NSF grades us on when they decide whether to fund a proposal or not, this one probably carries 90% of the weight and Broader Impacts carries about 10% of the weight." (R06)
Concern	Statements referring to concerns, hesitations, or pain points around the merit review criteria in theory, application, or practice. Could refer to BI/IM, or the proposal application and review process generally.	"And I guess my big concern with broader impacts can be that because they can feel frustrating for people, it's well, let's get a quick fix for this. Okay, here are the five things you list that have the most chance of success." (R03)
Institutional Support	Statements referring to institutional support of grant applications and the Merit Review criteria, particularly Broader Impacts. Can include critique of institutional support, recommendations for changes or improvements.	"And of the two criteria, the Intellectual Merit and Broader Impacts that NSF grades us on when they decide whether to fund a proposal or not, this one probably carries 90% of the weight and Broader Impacts carries about 10% of the weight." (R06)

Table 5. Modified Roberts (2009) coding criteria & definitions for Broader Impacts.

	Roberts (2009)	*Modified Code Definitions*
Criteria for Science		
Infrastructure for Science	Creation of new research methodology, tools, or data sources that will be useful to advance science	Creation of new research methodology, tools, or data sources that will be useful to advance science; **research outputs as broader impacts**.
Training and education	Includes mentoring undergraduates, graduate students, and postdoctoral fellows in the laboratory and teaching classes	Includes mentoring undergraduates, graduate students, and postdoctoral fellows in the laboratory and teaching classes, **as well as training and mentoring early career faculty**
Broadening participation	Recruiting or including under-represented groups in research or in outreach efforts. Includes efforts to attract women to science and to keep them in the academic pipeline for all fields **but excludes funding female students in biology and social sciences.**	Recruiting or including under-represented groups in research or in outreach efforts. Includes efforts to attract women to science and to keep them in the academic pipeline for all fields.
K-12 outreach	Outreach to kindergarten to 12th grade students or teachers helps to get kids excited about science and ensure a pipeline of future scientists.	Outreach to kindergarten to 12th grade students or teachers, helps to get kids excited about science and ensure a pipeline of future scientists. **Includes curriculum development and professional development for K-12 educators and administrators**
Academic collaboration	Research collaborations with other universities in the USA or abroad	Research collaborations with other universities in the USA or abroad, **intra-institutional collaboration across labs or research centers.**
Criteria for society		
Potential societal benefits	Direct claims that the research could help to inform policy, be useful for industry, or lead to some solution to a real-world problem; general statements of improved understanding of a natural or technical process (i.e. climate change or ecosystems) were not included	Direct claims that the research could help to inform policy, be useful for industry, or lead to some solution to a real-world problem; **includes claims that research would benefit society in broad or undefined ways.**
Outreach/ broad dissemination	Dissemination of research results for non-academic audiences in any form (web site, seminars, meetings, newspapers, magazines, etc). Does not include K-12 outreach projects	Dissemination of research results for non-academic audiences in any form (web site, seminars, meetings, **museum exhibitions**, newspapers, magazines, etc). Does not include K-12 outreach projects **or formal school curricula**.
Partnerships with potential users of research results	Includes partnerships with industry, non-profits, community groups, government bodies (local, state, national, or international), and national laboratories	Includes partnerships with industry, non-profits, community groups, government bodies (local, state, national, or international), and national laboratories

In addition to Deterding and Water's (2018) flexible coding approach, a modified version of Roberts (2009) coding scheme for types of Broader Impacts activities (Table 5) was also used to analyze survey and interview responses, specifically when survey respondents and interviewees were asked to discuss the types of Broader Impacts they had previously proposed, or wanted to see more of. That coding scheme, with the modified definitions used in this research project, is shown in Table 5. The two categories, Criteria for Science and Criteria for Society, represent Roberts' general groupings for the eight total types of Broader Impacts activities but were not used in the coding process itself.

Dedoose qualitative data analysis (QDA) software was used to facilitate coding. Intercoder reliability was not performed due to the small sample size of 10 researchers and 7 Program Officers interviewed, and exploratory nature of the research. With the author acting as the only primary coder, a focus on a detailed and nuanced understanding of individual codes was prioritized over generalizability. Instead, participant validation was used to check the accuracy of the coding process. Select codes and excerpts were shared with three participants (Two researchers and one Program Officer) for feedback. Validating participants were selected for existing familiarity with the qualitative coding process and code friction was settled via discourse.

4. Results and Discussion

The goal of this dissertation research was to deepen understanding of how individual researchers and program officers interpret the merit review process, with particular focus on the concept of weighting and its influence on perceptions of, and engagement with, Broader Impacts activities. Analysis of both survey and interview data revealed several key themes, particularly regarding the Merit Review criteria and the Broader Impacts criterion specifically.

4.1 Survey Results

The majority of survey participants were strongly supportive of Broader Impacts, as seen by the fact that 60% of respondents (N=40) indicated that including Broader Impacts in their research was “extremely important”, and 74% of respondents (N=49) strongly agreed or agreed that they were confident in their understanding of the Broader Impacts criterion.

However, despite indicating both positive valuation of and confidence in comprehension, 42% of respondents (N=28) also noted they found actually including Broader Impacts in their research to be “extremely, very, or moderately challenging”. A further 16% of total respondents (N=10), all of whom found including Broader Impacts to be challenging, agreed with the statement “including Broader Impacts takes away from time spent on my research”. This last result in particular indicates that Broader Impacts are still seen by some researchers as external or additional to the research process, potentially in a burdensome way.

Statistical analysis of the survey responses from those researchers who subsequently participated in an interview indicated no significant differences from the overall population’s average questions responses. While not significant, the greatest discrepancy between the responses of the interviewee subgroup from the overall group occurred with the statement “including Broader Impacts takes away from time spent on my research”, with the interviewees all answering “disagree” (36%) or “strongly disagree” (44%) compared to the overall population, where 19% agreed or were neutral (21%) on the statement. Tables 6 and 7 show researcher’s responses to Likert scale survey questions. Screening questions and long form responses are not shown and can be found in Appendix A.

Table 6. Likert scale survey responses.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	n
I am confident in my understanding of what the NSF Broader Impact criteria represents.	25	24	10	5	2	66
Including Broader Impacts in my research has enhanced the research itself	34	17	11	3	1	66
My research has benefitted from including Broader Impacts	34	16	13	2	1	66
My institution supports Broader Impacts research or activities	25	30	10	1	0	66
I have proposed Broader Impacts activities in my research but have been unable to execute such projects for any reason.	1	12	17	22	14	66
There are Broader Impacts activities I would like to incorporate into my research but have not been able to.	5	33	15	10	2	65
I find Broader Impacts requirements to be difficult to understand, vague, or imprecise	0	11	21	24	10	66
I have missed research opportunities because of Broader Impacts requirements	0	8	14	24	20	66
Including Broader Impacts considerations or activities does not enhance the value of my research	0	2	9	19	36	66
Including Broader Impacts considerations or activities takes away from time spent on my research	0	10	11	23	21	65

Table 7. Likert scale survey responses continued.

	Extremely Important	Very Important	Moderately Important	Slightly Important	Not At All Important	n
How important to you is including Broader Impacts in your research?	40	17	6	1	2	66
	Extremely Challenging	Very Challenging	Moderately Challenging	Slightly Challenging	Not At All Challenging	n
How challenging is including Broader Impacts activiites or considerations in your research?	2	8	18	15	23	66

4.2 Interview Results

The qualitative coding process highlighted several repeating motifs, with the primary codes Weighting, Funding, and Support emerging as strong themes. Select primary codes and definitions are shown in Table 4. As seen in Figure 2, in total, 10 Researchers responded to the emailed interview request and were interviewed. Separately, 17 current and former Program Officers were contacted about participating in an interview via personal networks and introductions from colleagues, mentors, and the dissertation committee. In total seven completed an interview. All seven were former Program Officers ranging from 3-10 years out of their posting and may be referred to as such during discussion of the results and findings.

Interview findings revealed key themes around the weighting of the two criteria, including concerns around the limitations of the Merit Review criteria, and a

perceived imbalance between the two criteria, with both groups of participants indicating an emphasis on Intellectual Merit over Broader Impacts in both the review process. and indirectly in career advancement.

4.2.1 Proposal Review

Both researchers and Program Officers reported a perceived imbalance, with Intellectual Merit often overshadowing Broader Impacts in the proposal review process. One researcher, when presented with the definition of the Intellectual Merit criteria during the artifact elicitation process, shared this:

> R06: "The Intellectual Merit criterion should be all about advancing knowledge. It should be about creating new knowledge, about creating new theory. That's the whole point. And of the two criteria, the Intellectual Merit and Broader Impacts that NSF grades us on when they decide whether to fund a proposal or not, this one probably carries 90% of the weight and Broader Impacts carries about 10% of the weight."
>
> DW: "And when you say the weight, what are you referring to?"
>
> R06: "Um, like if you slip up a little bit on Broader Impacts, but your Intellectual Merit is solid, you're in good shape. If it's the other way around, you are out of luck. "

Another researcher noted:

> R02: "You know, sure, broader impacts, it says you have to address it, so you have to address it, but nobody really pays attention to it."

From this and similar excerpts, the primary code "Weighting" emerged, representing participants' experiences with and perceptions of how each criteria is weighed and balanced against the other during the research, review, and post-evaluation processes. This theme of weight as a characteristic of the Merit Review Criteria and the NSF's review process echoes similar concerns raised during the proposal of the Intellectual Merit and criteria in 1997 (NSB-NSF Staff Merit Review Task Force). However, most participants described 'Weighting' as an externally

imposed expectation, evident in the review process and maintained and enforced within NSF policies and expectations, rather than as a reflection of their own values or research priorities. As one researcher explained,

> R03: “Um, I think intellectual merit weighs more highly. And part of that's just sort of like structural in the sense that I know that this is how proposals get reviewed, right. So like proposals don't get reviewed on, you know, how is the researcher maintaining a work life balance. Right? They get reviewed on the intellectual merit. So like I think just as a matter of tactics we would be compelled to focus on the intellectual merit, probably first and foremost.”

Despite maintaining a strategic alignment with perceived NSF expectations, R03 also expressed strong support for and personal conviction in the significance of Broader Impacts, stating later in their interview “to me, the broader impacts is conceptually the most important thing.” (R03). However, in the above excerpt, R03 notes feeling “compelled to focus on the Intellectual Merit”, indicating that their ‘tactical’ decision to heavily weight Intellectual Merit was not wholly in line with their own research goals, but rather a strategic decision made in response to perceived NSF expectations. This example, among others like it, highlights a tension between participants’ personal research values and the perceived demands of the NSF review process - a misalignment that may influence the direction of current and future research projects under current such evaluative norms.

Interviews with former Program Officers followed similar questions as for researchers (Appendix C), and were coded similarly as well. Among the former Program Officers interviewed, ‘Weighting’ Intellectual Merit and Broader Impacts also surfaced as a significant concern, mirroring statements made by researchers. During an interview, one former Program Officer related this interaction with a ‘lifer’, or long-term non-rotational Program Officer who had been with the NSF for many years:

> P03: “As a senior Program Officer, a lifer once said to me, Broader Impact doesn't matter if there's no Intellectual Merit. And so if you aren't actually advancing knowledge, it doesn't matter what the impact is.”

This perspective, shared by a former Program Officer in an engineering-related division, underscores a prevailing hierarchy that regularly places Intellectual Merit above Broader Impacts within the broader organizational culture of the NSF, and particularly within specific divisions as seen here. This, and other similar statements made by former Program Officers under the primary code Weighting, coupled with the fact that these concerns and tensions have endured so long, since the initial introduction of the criteria in 1997, indicates that the challenges related to evaluating the two criteria equally are not only a concern for applicants but also continue to resonate deeply with institutional officials, reflecting broader organizational tensions and priorities.

4.2.2 Tenure and Promotion

Researchers and former Program Officers also indicate that, despite internally recognizing the importance of Broader Impacts, they feel pressured to prioritize Intellectual Merit due to its greater influence on funding outcomes and institutional evaluations, such as tenure and promotion. This is reflected in the interview data, with statements indicating researchers perceive Broader Impacts as being valued secondary to more traditional metrics encapsulated within Intellectual Merit, like publications and successful grant numbers. This perception contrasts with survey results that show a high internal valuation of Broader Impacts among researchers. In short, despite researchers’ own internal valuation, external entities such as their own home institutions do not share that same level of value placed on Broader Impacts. The majority of researchers interviewed, (7 of 11, or 64%) indicated Broader Impacts was less valuable than Intellectual Merit related work for administrative or career

purposes at their home institution. Several reseasrchers interviewed specifically called out the promotion and tenure process, with one researcher stating:

> R09: "And so, when it comes down to it the Broader Impacts is still seen as the icing on the cake. When my tenure portfolio is reviewed, the number of federally funded grants, the number of publications that are scientific still holds more weight than community engagement…"

Similarly, former Program Officers interviewed also acknowledged a similar weighting of Intellectual Merit vs Broader Impacts in academic career progression specifically:

> P04: "You know, it is worth emphasizing that the training most faculty get is anti Broader Impact, right? I think they're enmeshed in a culture and that the reappointment promotion and tenure process, everything we do doesn't help in that."

The quote from the former Program Officer shared above emphasizes a view that the relationship between faculty training and the research cultures that shape and influence their work—particularly within the context of the promotion and tenure process—is, at best, not 'helpful', and may even be devaluing or antagonistic toward Broader Impacts. What the researchers experience in their own individual cases, the former Program Officer alludes to as a larger, more systemic concern. Given the significant role tenure plays in academic career trajectories, and the role of research and publication in attaining tenure, it is yet more important to understand how the National Science Foundation's merit review criteria are implemented in practice, especially concerning their influence on academic career trajectories.

4.3 Weighting as 'Good Science':

Both discussions of weighting Intellectual Merit versus Broader Impacts, and such weighting's downstream effects on researcher's career trajectories emphasize ways in which external tensions around the Merit Review criteria have come to shape the landscape of scientific funding and the processes of research in practice.

However, researchers and Program Officers also both indicated internal value structures and tensions around if both criteria can or should be equally weighted or expected of researchers. Among former Program Officers interviewed, one noted the following stance on balancing Intellectual Merit and Broader Impacts:

> P03: "And so then the balance needs to be whatever gives you a clear enough understanding of the Intellectual Merit that you can see the potential to benefit society. If you can't name that, if your methods aren't supported, if your foundation is sandy then it's not going to matter what you come up with, because I don't want you spreading around bad knowledge."

Additionally, concerns about asking scientists to be good at all of these things (aka. Doing rigorous research and also public outreach) emerged:

> R02: "Why are we asking scientists and NSF to pay for a dysfunctional K-12 system [...]? That should not be a scientist's job. Finding people who are great at all of these things narrows the field and potentially leaves us without."

When asked to elaborate, this researcher followed up by explaining:

> R02: "So like in that way, one of the things that I worry about with Intellectual Merit and Broader Impacts, if you've got a researcher who is exceptionally good at broader impacts - they are engaging. They have a really wide social media platform. They translate complicated ideas into approachable things. They work really well with policymakers, but their experimental design is like, eh, yeah, okay. Like it's not really pushing the field forward. How does NSF choose to value that proposal?
>
> And on the other side, if you've got someone who's blatantly Asperger's, right. and is only interested in progressing the science, but God, are they good at it, right? And they've got just novel ideas, but like, they're not good at the broader impacts. Like we might need to have just 2 or 3 that like, yes... at least they have postdocs and grad students in their broader impacts? Just let them go. They're going to do something amazing that can be really tricky. And I think that that's something that has to be handled at the program manager level. Right? Like if something is so brilliant, that honestly, what that person should do is partner with a broader impacts person."

These perceptions underscore challenges within both the NSF review process and the practice and culture around research impact at academic institutions, whereby traditional academic achievements are more heavily valued than societal contributions.

5. Limitations and Generalizability

Respondent bias poses a notable limitation when exploring perceptions of the NSF's merit review criteria. Particularly, social desirability bias may skew the data, as the participant researchers and former Program Officers, might prefer to portray themselves in a favorable light by not expressing negative views about the Broader Impacts criterion due to their professional roles and the potential repercussions of voicing controversial opinions in a professional setting. Additionally, self-selection bias poses another limitation, whereby individuals with strong opinions about the NSF criteria could reasonably be expected to be more likely to participate in the survey or interviews, potentially leading to an overrepresentation of extreme or stronger views.

Together, these biases could limit the generalizability of the findings and obscure the nuanced perspectives that might be more representative of the broader academic community. The generalizability of this exploratory research project is further constrained by its limited and homogeneous sample size, consisting of interviews with only 11 researchers and 7 former Program Officers. The use of non-generalizable, exploratory results from qualitative coding is justified given the focus on capturing the nuanced, context-specific perspectives of individual respondents, which is particularly valuable for gaining in-depth insights from a small, purposefully selected sample where saturation can be reasonably achieved within 9 interviews (Hennick and Kaiser 2022). However, this small research population, coupled with the fact that all surveyed and interviewed researchers were drawn from a single research university in a single-site study, limits the extent to which the findings can be applied to the broader academic population or other institutional contexts. As such, the insights gained, while valuable, must be interpreted with caution when considering their applicability to different settings or a wider range of academic perspectives.

6. Conclusion:

This research sought to provide a nuanced understanding of the National Science Foundation's Broader Impacts criterion and its interaction with the Intellectual Merit criterion in the current research landscape, revealing both new and enduring questions and concerns around Broader Impacts within the research community. Contrary to some earlier literature showing widespread resistance (National Science Foundation 2011), a substantial proportion of survey participants (60%, N=40) recognized the significance of Broader Impacts in their research. Moreover, a majority (74%, N=49) expressed confidence in their comprehension of the criteria, suggesting that previous efforts to clarify these expectations have been somewhat effective.

However, the practical incorporation of Broader Impacts into research remains challenging for a significant segment of the scientific community, with nearly half of survey respondents finding the integration of Broader Impacts into their research to be challenging, and some researchers stated that Broader Impacts activities detracted from their primary research efforts. These perceptions highlight a persistent tension where Broader Impacts is still viewed by some as a supplementary rather than an inherent component of scientific research, potentially undermining its intended impact.

Further complicating the landscape, the study reveals a perceived imbalance in the weight given to Intellectual Merit and Broader Impacts during the proposal review process. This imbalance, identified as a primary theme of 'Weighting' in qualitative analysis, suggests that despite the formal mandate of ‘equal consideration’ of the two criteria, Intellectual Merit often overshadows Broader Impacts in practice. Both researchers and Program Officers report feeling compelled to prioritize Intellectual Merit due to its perceived greater influence not only funding decisions

but also career advancements, such as tenure and promotion. This discrepancy is not only observed by researchers but also reflected in the testimonies of former Program Officers, indicating that it is a pervasive issue both individually and within the greater organizational culture of the NSF. The enduring nature of these challenges, first noted during the initial introduction of the criteria in 1997 (NSB-NSF Staff Task Force on Merit Review), points towards an ongoing critical gap between policy intentions and practical implementation.

Despite the broad familiarity with and approval of Broader Impacts noted in this study, researchers are nevertheless struggling with launching, supporting, and validating the inclusion of Broader Impacts work in their research. As such, there is a clear need for the NSF and academic institutions to enhance guidance and support, helping researchers to effectively integrate and realize both criteria. This study advocates for more equitable consideration and reward for efforts related to both Intellectual Merit and Broader Impacts, suggesting that merely acknowledging their importance is insufficient without corresponding action. At the institutional level, adjustments in the review process to genuinely balance the weight given to Intellectual Merit and Broader Impacts can encourage researchers to engage more deeply with the societal impacts of their work.

To more effectively foster an environment where Broader Impacts is both valued in theory and empowered in practice, it may be beneficial for the NSF and institutions supporting researchers to consider revisiting and strengthening the mechanisms for integrating these criteria into the research evaluation and career processes. Moving forward, it is crucial for the NSF and the broader scientific community to address these disparities and work towards a more balanced approach that genuinely reflects the dual objectives of advancing scientific knowledge and contributing to societal needs.

7. References

Ambrose, M. L., & Cropanzano, R. (2003). A longitudinal analysis of organizational fairness: An examination of reactions to tenure and promotion decisions. Journal of Applied Psychology, 88(2), 266.

Bouwma-Gearhart, J., Lenhart, C., Carter, R., Mundorff, K., Cho, H., & Knoch, J. (2021). Inclusively recognizing faculty innovation and entrepreneurship impact within promotion and tenure considerations. Journal of Open Innovation: Technology, Market, and Complexity, 7(3), 182.

Bozeman, B., & Boardman, C. (2009). Broad impacts and narrow perspectives: Passing the buck on science and social impacts. *Social Epistemology*, *23*(3-4), 183-198.

Chalmers, D. (2011). Progress and challenges to the recognition and reward of the scholarship of teaching in higher education. Higher Education Research & Development, 30(1), 25-38.

Doberneck, D. M. (2022). Are we there yet?: Outreach and engagement in the consortium for institutional cooperation promotion and tenure policies. Journal of Community Engagement and Scholarship, 9(1).

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. PloS one, 5(4), e10271.

Frodeman, R., & Holbrook, J. B. (2012). The promise and perils of transformative research. *UNT Digital Library. http://digital. library. unt. edu/ark:/67531/metadc84363*.

Frodeman, R., Holbrook, J. B., Bourexis, P. S., Cook, S. B., Diederick, L., & Tankersley, R. A. (2013). Broader Impacts 2.0: seeing—and Seizing—the opportunity. BioScience, 63(3), 153-154.

Gusterson, H. (1997). Studying up revisited. PoLAR, 20, 114.

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine*, *292*, 114523.

Holbrook, A. L., Krosnick, J. A., & Pfent, A. (2007). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. Advances in telephone survey methodology, 499-528.

Kamenetzky, J. R. (2013). Opportunities for impact: Statistical analysis of the National Science Foundation's broader impacts criterion. Science and Public Policy, 40(1), 72-84.

Lok, C. (2010). Science for the masses: the US national science foundation's insistence that every research project addresses' broader impacts' leaves many researchers baffled. Corie Lok takes a looks at the system. *Nature*, *465*(7297), 416-419.

McCullough, J. (1989). First comprehensive survey of NSF applicants focuses on their concerns about proposal review. Science, Technology, & Human Values, 14(1), 78-88.

Nagy, D. (2016). Determinants of Broader Impacts Activities: A Survey of NSF-Funded Investigators. Journal of Research Administration, 47(1), 68-93.

NSB-NSF Staff Merit Review Task Force. (1997, March). Merit Review Final Recommendations. https://www.nsf.gov/pubs/1997/nsbmr975/nsbmr975.htm

NSB Task force on Merit Review. (2011, December). *National Science Foundation's Merit Review Criteria: Review and Revisions*. National Science Board. https://www.nsf.gov/nsb/publications/2011/meritreviewcriteria.pdf

National Science Foundation (2002). Important Notice No. 127 (in127). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in127.pdf

National Science Foundation. (2005). Report of the National Science Board on the National Science Foundation's merit review system. Retrieved from www.nsf.gov/nsb/documents/2005/0930/merit_review.pdf

National Science Foundation. (2012, October 4). *NSF Revises Merit Review Criteria.* https://www.nsf.gov/bfa/dias/policy/merit_review/factsheets/proposers.pdf

National Science Foundation (2013). *Proposal and Award Policies and Procedures Guide* (NSF 13-1, NSF13001). Washington, DC. Accessed 13 June 2024 at https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/nsf13_1.pdf

National Science Foundation (2021). *Proposal and Award Policies and Procedures Guide* (NSF 22-1, OMB Control Number 3145-0058). Washington, DC.

National Science Foundation (2023). *Proposal and Award Policies and Procedures Guide* (NSF 24-1, NSF24001). Washington, DC. Accessed 22 June 2024 at https://nsf-gov-resources.nsf.gov/files/nsf24_1.pdf

Oleschuk, M. (2020). Gender equity considerations for tenure and promotion during COVID-19. Canadian review of sociology, 57(3), 502.

O'meara, K. (2010). Rewarding multiple forms of scholarship: Promotion and tenure. Handbook of engaged scholarship: Contemporary landscapes, future directions, 1, 271-294.

Perna, L. W. (2001). Sex and race differences in faculty tenure and promotion. Research in higher education, 42, 541-567.

Roberts, M. R. (2009). Realizing societal benefit from academic research: Analysis of the National Science Foundation's broader impacts criterion. Social Epistemology, 23(3-4), 199-219.

Rothenberg, M. (2010). Making judgments about grant proposals: A brief history of the merit review criteria at the National Science Foundation. *Technology & Innovation*, *12*(3), 189-195.

Salet, W. (2018). *Public norms and aspirations: The turn to institutions in action*. Routledge.

Sarewitz, D. (2011). The dubious benefits of broader impact. *Nature*, *475*(7355), 141-141.

Schimanski, L. A., & Alperin, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. F1000Research, 7.

Tijdink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication pressure and scientific misconduct in medical scientists. Journal of Empirical Research on Human Research Ethics, 9(5), 64-71.

Van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. Journal of the American Society for Information Science and Technology, 63(7), 1282-1293.

Watts, S. M., George, M. D., & Levey, D. J. (2015). Achieving broader impacts in the national science foundation, division of environmental biology. BioScience, 65(4), 397-407.

Wiley, S. L. (2014). Doing broader impacts? the National Science Foundation (NSF) broader impacts criterion and communication-based activities.

Winkler, J. A. (2000). Faculty reappointment, tenure, and promotion: Barriers for women. The Professional Geographer, 52(4), 737-750.

Woodson, T., & Boutilier, S. (2023). From intent to impact—The decline of broader impacts throughout an NSF project life cycle. *Research Evaluation*, *32*(2), 348-355.

CHAPTER 3

BROADENING THE SCOPE OF BROADER IMPACTS

Broadening the Scope of Broader Impacts: Evolving Guidelines, Partial Success and Ongoing Barriers

1. Introduction

The National Science Foundation (NSF) has been at the forefront of scientific funding and research since its establishment in 1955. Furthermore, since its founding the NSF has also maintained an expectation that the public impact of the research it funds must be considered both in the application and internally when making funding decisions, even though that expectation has not always been overt (Rothenberg 2010, Center for Advancing Research Impact in Society 2023a).

Since revising its Merit Review guidelines in 1997 to include the Broader Impacts (BI) criterion, explicitly bring research impact to the foreground for the first time, the NSF has continuously sought to clarify and refine its expectations with the goal of not only enhancing the immediate academic rigor of proposals but also fostering a wider application of scientific research for societal benefit (Rothenberg 2010, NSB-NSF Staff Merit Review Task Force 1997). However, the journey toward a comprehensive and universally understood Broader Impacts criterion has been fraught with challenges, characterized by evolving guidelines (MacFadden 2019, Center for Advancing Research Impact in Society 2023b), community uncertainties (Mervis 2011), and myriad interpretations of what constitutes meaningful societal impact - along with the kinds of activities that might suffice (Wiley 2014, Nadkarni and Stasch 2013).

NSF's efforts to embed a culture of Broader Impacts began in earnest following the 1997 guideline revisions, with subsequent Important Notices and other

missives directly addressing the academic community's requests for clearer definitions and examples of acceptable Broader Impacts activities (NSF 1997, 1999, 2002, 2007a). Yet, the built-in flexibility intended to stimulate innovative applications of research for public good often muddied the waters further, leading to criticisms of vagueness and the unintended formation of a de facto checklist mentality (Mervis 2011).

Furthermore, the academic community's response to the Broader Impacts criterion has historically involved a mix of enthusiasm and skepticism, with some of the original concerns and critiques around weighting and the rewards (or lack thereof) researchers receive for doing Broader Impacts work discussed in Chapter 2. As a result, universities and research institutions have ramped up support for researchers through various means, from hosting Broader Impacts resource pages to establishing dedicated centers, which offer specialized services ranging from project design to evaluation. However, the question remains: Are these efforts translating into genuine societal impacts, or are they simply ticking boxes to satisfy funding requirements? And even beyond that remains a larger question: What, ultimately, is the goal of the Broader Impacts criterion - better research impact, or better researchers?

Following this question, this paper focuses on examining the types of Broader Impacts activities both presented by the National Science Foundation and performed by researchers. It also considers the frustrations and challenges associated with these activities, and the various resources, centers, and messages that have developed around the 'doing' of Broader Impacts in order to better understand how Broader Impacts has been integrated into the research lifecycle - and researchers' professional identities (Risen and Storksdieck, 2018) since the introduction of the criterion over two decades past in 1997.

2. The Breadth of Broader Impacts: Types and Centers

2.1 Evolving Guidelines and Expectations for Broader Impacts

During the introduction of the revamped Merit Review criteria in 1997, concerns about the vagueness of the Broader Impacts criterion were widespread, as seen during the public comment period (NSB-NSF Staff Merit Review Task Force, 1997); follow-up statements by the NSF in Important Notices 121 (1997), 125 (1999), and 127 (2002) showcased a increasing effort on the part of the institution to overcome resistance (Mervis, 2007), and support and guide members towards both valuing and implementing Broader Impacts in their research.

In Important Notice 125, for example, the NSF reached out to the Presidents of Universities and Colleges, saying "We want to ensure that the criterion relating to Broader Impacts is considered and addressed in the preparation and review of proposals submitted to NSF. We encourage you to emphasize within your organization the importance of both criteria -- to principal investigators who are developing and submitting proposals as well as to those who may serve as proposal reviewers." (1999). This notice indicates that concerns about the uptake of Broader Impacts in the research community still lingered, and were substantial enough to necessitate an Important Notice addressing those concerns. Likewise, in 2002, Important Notice 127 informed the community that the NSF would "return without review proposals that do not separately address both merit review criteria within the Project Summary"

Yet, confusion persisted, resulting in the first major re-evaluation of the Merit Review criteria in 2007, broadcast to the research community via a Dear Colleague Letter (NSF 2007a). Table 8 shows the evolution of possible kinds of Broader Impacts activities listed in first the Dear Colleague Letter (NSF, 2007a), and then in subsequent editions of the Proposal and Award Policies and Procedures Guide, aka the PAPPG, from 2007-2020 (NSF, 2007b; 2013; 2020).

Table 8. Evolution of possible Broader Impacts activities listed in official NSF documentation.

2007	2007	2013	2020
Dear Colleague Letter	PAPPG 13-1 II.C.2.d(i)	PAPPG 13-1 II.C.2.d(i)	PAPPG 20-1 II.C.2.d(i)
Advance discovery and understanding while promoting teaching, training, and learning	How the project will integrate research and education by advancing discovery and understanding while at the same time promoting teaching, training, and learning	Full participation of women, persons with disabilities, and underrepresented minorities in science, technology, engineering, and mathematics (STEM)	Full participation of women, persons with disabilities, and underrepresented minorities in science, technology, engineering, and mathematics (STEM)
Broaden participation of under-represented groups	Ways in which the proposed activity will broaden the participation of underrepresented groups (e.g., gender, ethnicity, disability, geographic, etc.)	Improved STEM education and educator development at any level	Improved STEM education and educator development at any level
Enhance infrastructure for research and education	How the project will enhance the infrastructure for research and/or education, such as facilities, instrumentation, networks, and partnerships	Increased public scientific literacy and public engagement with science and technology	Increased public scientific literacy and public engagement with science and technology
Broaden dissemination to enhance scientific and technological understanding	How the results of the project will be disseminated broadly to enhance scientific and technological understanding	Improved well-being of individuals in society	Improved well-being of individuals in society
Benefits to society	Potential benefits of the proposed activity to society at large	Development of a diverse, globally competitive STEM workforce	Development of a diverse, globally competitive STEM workforce
n/a	n/a	Increased partnerships between academia, industry, and others	Increased partnerships between academia, industry, and others
n/a	n/a	Improved national security	Improved national security
n/a	n/a	Increased economic competitiveness of the United States	Increased economic competitiveness of the United States
n/a	n/a	Enhanced infrastructure for research and education	Enhanced infrastructure for research and education
n/a	n/a	n/a	Use of science and technology to inform public policy

The 2007 Dear Colleague letter began by reminding readers of the two Merit Review criteria, Intellectual Merit and Broader Impacts, and the guiding questions that frame both criteria in the NSF Grant Proposal Guide, then went on to note:

> "It may be helpful to illustrate the kinds of activity that are appropriate to each of the questions above. Caveat lector - the following list is neither prescriptive nor exhaustive and should not be read in ways that constrain the creativity of researchers in proposing activities with broader impact."

What followed was the first of several attempts by the National Science Foundation to provide a list of potential activities that would suit the Broader Impacts criterion (MacFadden, 2019). The above text in Table 8 is taken from the Project Description section of the respective editions of the PAPPG. This section explicitly instructs potential grant applicants to include a distinct Broader Impacts section. This requirement is in addition to other elements of the project description detailing the general plan of work, objectives, and other technical contributions of the proposed research. Importantly, the 2020 edition of PAPPG, concludes with a new additional caveat: "These examples of societally relevant outcomes should not be considered either comprehensive or prescriptive." (NSF, 2020, II.D.2.d(i))

As such, potential applicants are informed that this list of outcomes is not to be taken as a checklist, emphasizing that the listed outcomes are merely possible options, and should not be perceived as an all-encompassing list of possibilities. Additional resources often highlight this fact, as seen with the Broader Impacts resource page from the University of Wisconsin-Madison, which notes that Broader Impacts "are designed to be broad so that scientists have the flexibility and freedom to design their Broader Impacts around their interests and complements their research." (2024). The goal here, in theory, is to allow applicants the flexibility to construct Broader Impacts that are well tailored to their research, and to encourage applicants to think beyond the provided examples and consider other innovative

ways their research could benefit society or advance scientific knowledge (MacFadden, 2019).

In practice, that offer of flexibility has not always played out as intended. After the introduction of the initial 2007 list (NSF, 2007a), seen in Table 1, community backlash to the listed activities ranged from disagreement with specific listed activities, to wholesale contestation of the inclusion of Broader Impacts at all (Mervis 2011). What the National Science Foundation saw as flexibility, others often interpreted as a lack of clarity (MafcFadden 2019, Mervis 2011, Rothenburg 2010).

The Center for Advancing Research in Society's (ARIS) Broader Impacts Toolkit encourages researchers to consider their 'Impact Identity' as an evolving and multi-faceted part of their research career when developing Broader Impacts projects (Risen and Storksdieck, 2018) and in various places encourages applicants to be 'creative' and innovative in their proposed Broader Impacts activities (Center for Advancing Research in Society, 2023c). However, the first question of their Broader Impacts Plan Checklist, included in the Toolkit, also asks researchers to check off whether or not the proposed project "addresses one of the target outcomes for Broader Impacts activities outlined by NSF", and proceeds to list the same options specified in Table 1 under the 2020 edition of the PAPPG (NSF, 2020). This effectively re-positions the possible activities presented by the NSF as the very checklist they hoped to move away from (MacFadden, 2019).

2.2 Broader Impacts Types and Activities

While the National Science Foundation's guidelines continually evolve, and associated resources follow, encouraging flexibility and a broad array of potential activities in theory has shown varying degrees of implementation in practice. Extensive literature has explored the diversity of Broader Impacts activities proposed within these applications (Roberts, 2009; Wiley, 2014; Nagy, 2016). Fewer studies

have investigated activities actually conducted after proposal, however Woodson and Boutilier (2023) observed that the number of Broader Impacts activities initially proposed in NSF grant applications tends to exceed those actually implemented, with a significant decrease observed as projects progress.

In researching the variability of Broader Impacts activities proposed by researchers, Roberts (2009) developed a dual-category coding scheme for Broader Impacts activities. The first category, Criteria for Science, encompasses activities designed to enhance the infrastructure of scientific research, academic training and educational initiatives, efforts to broaden participation, outreach to K-12 education sectors, and projects fostering academic collaborations. Roberts' second category, Criteria for Society, includes activities aimed at achieving direct societal benefits, such as broad dissemination and public outreach efforts, and developing partnerships with community or industry stakeholders. These categories are themselves based off those initially presented by the NSF in their 2007 Dear Colleague Letter that first described a list of potential types of Broader Impacts researchers could include in proposals (NSF, 2007a).

Building off these categories, additional research indicates that the most frequently proposed Broader Impacts activities are those focused on education and training, such as conference presentations and undergraduate student training, which predominantly facilitate peer-to-peer engagement within the academic community (Wiley 2014; Kamenetzky 2013; Roberts 2009). Activities aimed at disseminating findings to non-academic audiences, particularly K-12 groups, also feature prominently, reflecting a consistent trend across both specific NSF divisions and the broader agency (Nadkarni and Stasch, 2013; Watts et al., 2015).

2.3 Broader Impact Centers and Professionalization

As the range of Broader Impacts activities thought up by the research community expands, and as uptake of the Broader Impacts criterion as an important aspect of research grows, so does a community of practitioners and professionals dedicated to supporting the development of Broader Impacts initiatives. The largest and most well known, ARIS emphasizes its specific focus on the NSF's broader Impact criteria: "The resources and tools on this site are designed to help Researchers and BI Professionals develop projects and partnerships that will satisfy the Broader Impact requirement of National Science Foundation (NSF) proposals, and help you fulfill your interest in communicating your science." (Center for Advancing Research in Society, 2023c).

Similarly, most research-focused universities offer at least a webpage of resources about Broader Impacts for researchers in the grant-writing process. A search of the top 50 national universities as ranked by US News shows that out of 53 institutions, 39 (74%) have an institutional resource page on Broader Impacts. The resources provided range from simple links back to NSF materials, to training materials on Broader Impacts, all the way to extensive intra-institutional networks of 'BI partners' and activities and editing and evaluation services for Broader Impacts proposals.

Other institutions of higher education offer dedicated centers and institutes focused on supporting Broader Impacts or offering Broader Impacts services for researchers and projects within their campuses, including the University of Notre Dame's Center for Broader Impacts (Notre Dame, 2024), Princeton's Council on Science and Technology (Princeton, 2024), and Virginia Tech's Center for Educational Networks and Impacts (Virginia Tech, 2024). Services offered range from assistance with design and development, to existing structured activities available for 'plugging in' (Notre Dame, 2024). Virginia Tech's center in particular

offers potential ‘BI Partners’ a menu of Broader Impacts services from Broader Impacts design reviews to full evaluation plans; services are provided on a funding-based ‘sliding scale’ starting at $500 for a field trip to a budget line-item of $35,000 to develop and maintain a museum exhibit (2024).

2.4 Current State of Broader Impacts

In asking researchers to contribute to ‘societal benefit’ through the Broader Impacts criterion, the National Science Foundation asked researchers to consider not only the immediate outcomes of their project for themselves and their field, but also how that work might extend beyond the individual researcher to society at large. Since the term "societal benefit" is itself conceptually diffuse, the addition of supplemental guidelines and lists of potential kinds of activities that might suit the Broader Impacts component of a particular research project unsurprisingly quickly followed, as shown in Table 1.

Yet, given the evolution of the Broader Impacts criterion since its introduction in 1997, and the increasing emphasis placed on requiring researchers and universities to meet the criteria in order to obtain funding from the NSF, it is understandable that even the supplemental guidelines themselves led to the development of a large body of external resources and professional organizations devoted to Broader Impacts. The growing trend of establishing dedicated Broader Impacts centers devoted to designing and executing Broader Impacts activities, either in collaboration with researchers or even all the way through to the complete planning, execution, and management of these projects, points towards a new approach to Broader Impacts work that prioritizes the funding-to-impact outcome over researcher integration and involvement.

Such centers may provide needed partnership and support to early-career researchers or those with minimal experience in NSF grant processes, public-facing

impact work, or limited community networks. However, it is also possible that more intensive management of Broader Impacts projects by some centers may distance researchers from impact work entirely. Understanding how various individual researchers and former Program Officers have approached the breadth of potential Broader Impacts types in their own research or divisions respectively, and how they perceive the emerging trend towards Broader Impacts centers, may illuminate in greater detail how Broader Impacts work is practically functionalized in actual research work.

3. Methodology

3.1 Study Design

The primary research question addressed in this chapter was "How and why do stakeholders engaging with the NSF - either as researchers, former Program Officers, or institutional officials - construct and value research impact, particularly as a function of the NSF's Broader Impact criteria?" And specifically, "How and why do individual scholars understand and value 'Broader Impacts' as an aspect of their own research and experiences?". Here, the two primary participant populations of individual researchers who have submitted NSF grants and former Program Officers were chosen for their particular roles within the research lifecyle. Researchers were a crucial focus of this dissertation because they are the primary actors responsible for interpreting and implementing the National Science Foundation's merit review criteria in their work, navigating the practical challenges these requirements present. Conversely, former Program Officers' perspectives on the merit review criteria were equally important, as they hold a unique responsibility for managing and maintaining a portfolio of research projects. Their role positions them at the intersection of individual and institutional perspectives, allowing them to mediate between the practical challenges faced by researchers and the broader strategic goals of the

institution. The data discussed in this chapter offers a particular focus on the types of Broader Impacts activities presented by the National Science Foundation and how those are done in practice by researchers, as well as exploring the emerging development of Broader Impacts focused resources and centers, to better understand how Broader Impacts have been understood by researchers and Program Officers alike.

3.2 Survey

A mixed methods survey, involving both short- and long-form questions that allowed researchers to elaborate on their responses in their own words, was disseminated to better understand researchers' familiarity with, perceptions of, and experiences with respect to the National Science Foundation's Merit Review criteria, with particular emphasis on the Broader Impacts criterion. As described in Chapter 2, participants were identified from an interdisciplinary scholarly network at a large public R1 university, which focuses on addressing global issues; participants were generally expected to be both familiar with and potentially strongly positively inclined towards Broader Impacts as a result, yielding a survey population with potential pre-existing familiarity with the criteria. Recruitment was facilitated through the network's official mailing list. The likely pro-Broader Impacts alignment of network members was recognized; however as little literature has directly addressed concerns among those who are still ultimately proponents of Broader Impacts, the survey population was still considered worthwhile in this instance.

The survey was administered digitally via the QuestionPro secure online survey platform, and was comprised of 25 questions designed to elicit both quantitative and qualitative responses. This included 14 short-form Likert scale questions to quantify attitudes and perceptions, alongside 11 open-ended questions intended to capture detailed, narrative data on participants' experiences and

viewpoints (Appendix A). The survey was intended to be completed in 30 minutes or less, a duration confirmed via pre-dissemination testing with two pilot testers from the target population, who were then excluded from participation in the final survey. To ensure anonymity and minimize response bias, surveys were conducted anonymously, and the survey protocol attained IRB approval (Appendix E). Upon completion, participants were given the option to provide their contact information voluntarily and offered the opportunity to engage in a more in-depth follow-up interview.

3.3 Semi-Structured Interviews

Semi-structured interviews were conducted to explore in detail the perspectives of both researchers and Program Officers officers regarding Broader Impacts, and the Merit Review criteria more generally. Following the survey phase, researchers who had opted to provide their contact information were invited for interviews via their supplied email addresses. Similarly, current and former Program Officers were approached through personal and professional networks for voluntary participation in the study.

The interview protocol was standardized across both participant groups to maintain consistency in data collection, although questions were minimally modified as needed to reflect the specific roles of the participants, either as researchers or Program Officers (See Appendices B and C). To ensure the effectiveness and clarity of the interview questions, the protocol was pilot tested with two individuals from the respective groups before the commencement of the full interview data collection phase, and received IRB approval (Appendix E).

All interviews were carried out over a fixed period from January to March 2024. This approach ensured that data collection was systematically organized and that all participants were engaged in a similar manner and time frame, allowing for

comparability across the two distinct groups while accommodating their unique perspectives within the NSF's review process. Audio recordings from Zoom interviews were labeled with anonymized participant IDs and securely stored in a designated cloud storage site; recordings were then transcribed using a denaturalized approach that focused on content rather than speech mechanics (Oliver et al., 2005). Cleaned transcriptions were uploaded to Dedoose for systematic qualitative coding and analysis.

3.3 Artifact Elicitation

In this study, the artifact elicitation method was employed during interviews with researchers and Program Officers to facilitate a deeper understanding of the NSF's Merit Review criteria, as was described during Chapter 2. Using Microsoft PowerPoint, two virtual text artifacts were created to define each criterion (Figures 6 and 7)

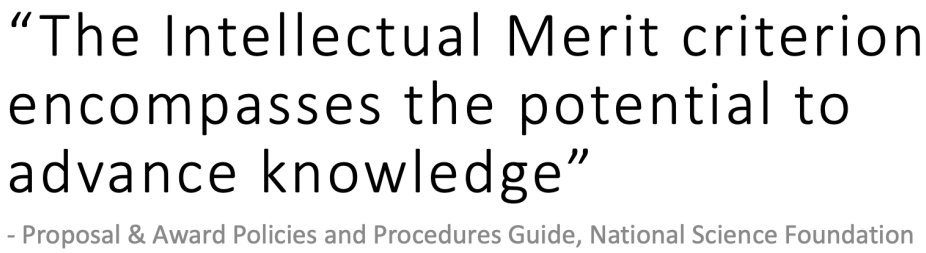

Figure 6. Intellectual Merit criterion artifact presented to all interview participants.

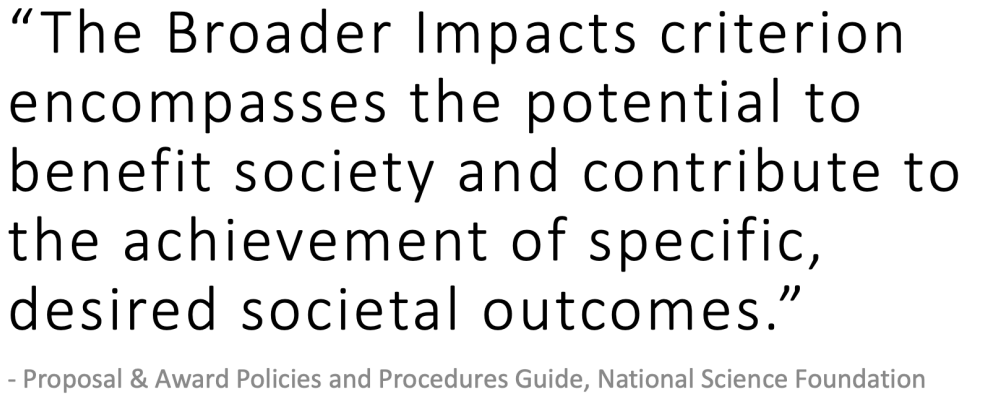

Figure 7. Broader Impacts criterion artifact presented to all interview participants.

Participants were initially presented with the Intellectual Merit slide and prompted to articulate their understanding of and personal experience with Intellectual Merit; participants were also asked to discuss whether their interpretation aligned with that of their colleagues or the NSF’s expectations. This process was subsequently replicated with the Broader Impacts slide; both the research and Program Officer participant groups followed the same procedures.

These visual artifacts served as critical universal central reference points during the interviews, enabling participants to delve into detailed discussions and reflections about both Broader Impacts and Intellectual Merit. The structured presentation was intended to standardize the information each participant received and provided participants a concrete touchstone to refer to as they explored their thoughts, while also allowing for flexibility in participants’ reactions and responses.

3.4 Qualitative Coding

For the analysis of interview data in this study, the "Flexible Coding of In-depth Interviews: A Twenty-first-century Approach" by Deterding and Waters (2018) was employed, in conjunction with a modified Broader Impacts activities coding

scheme developed by Roberts (2009). Flexible coding is particularly suited to scenarios where the complexities of the data emerge progressively, allowing for ongoing adjustments to the coding scheme. This flexibility ensures that the analysis remains comprehensive and aligned with emerging insights, an essential feature given the intricate nature of the topics discussed by participants who are typically more accustomed to conducting interviews than being interviewed.

This role reversal also introduces specific challenges associated with 'studying up' (Gustarson, 1997); documenting and understanding these dynamics was crucial for a thorough analysis while acknowledging the influence of hierarchical academic structures on the data interpretation. The stages of the flexible coding method by Deterding and Waters (2018) as applied to this research project are noted in Figure 8.

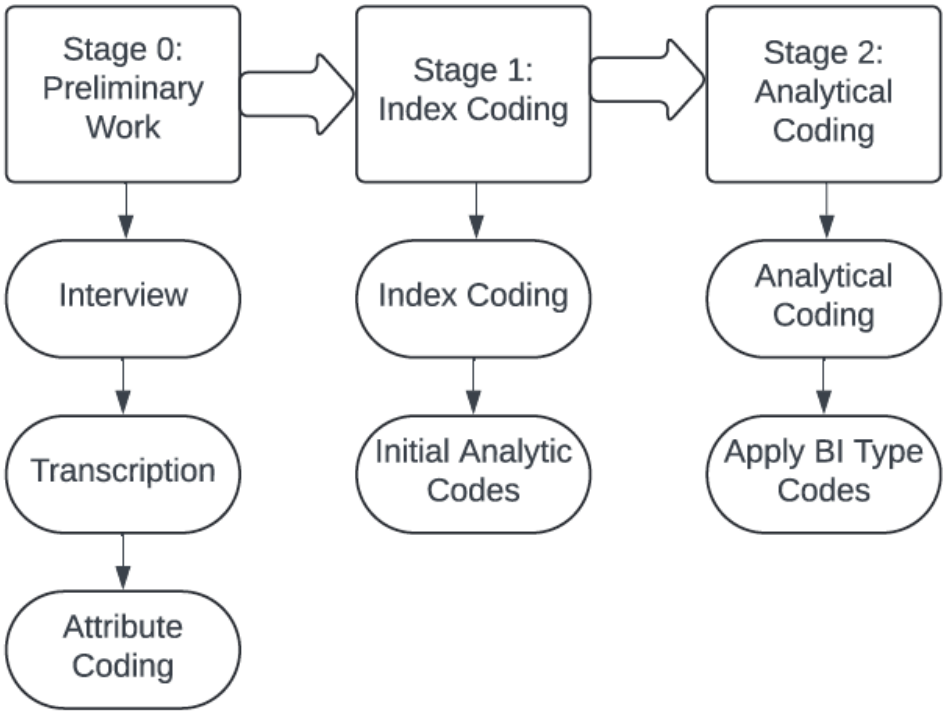

Figure 8. Stages of Deterding and Waters (2018) flexible coding approach.

Attribute codes were applied to each document, and included the participant's anonymized ID, participant group, and self-identified general research field or discipline. Index codes were then applied on the basis of the research questions and the interview questions, with each response to an interview question index coded as one unit in its entirety. Lastly, analytical codes were applied on a more fine-grained basis, using both in vivo codes representing participants actual phrasing, as well as an existing coding scheme for Broader Impacts types (Roberts, 2009). Attribute codes and select index and analytical codes are shown in Table 9.

A modified version of Roberts (2009) 'BI coding scheme', itself based on NSF guidance on Broader Impacts (2007) was also used to code survey and interview responses, in order to facilitate comparison of Broader Impacts activities discussed by participants with existing analyses of Broader Impacts activity variation. Following the NSF document from which Roberts developed these criteria, the eight types of Broader Impacts activities are divided into two larger categories: Criteria for Science, and Criteria for Society. These categories were created by Roberts and reflect general groupings of the eight types. This scheme was applied with minor changes during the coding process, and modifications to the original scheme are shown in Table 10 in bold, compared to Roberts' (2009) original definitions. Modifications were made to more accurately reflect current NSF clarifications on each type provided. These modifications were added to reflect the evolution of the suggested criteria types provided by the NSF as previously shown in Table 8.

Table 9. Example Index and Analytical codes as applied.

Code Type	Application or Definition	Example
Index Codes		
Understandings of BI	Interview Question, asked during Artifact Elicitation with Broader Impacts criterion definition displayed: "Personally, how do you understand Broader Impacts in your own research?"	"But broader impacts for me is a range of different societal benefits, whether it is education of the general populace, engagement with the general populace helping people understand what science is and that what is truth?" (R03)
Understandings of IM	Interview Question, asked during Artifact Elicitation with Intellectual Merit criterion definition displayed: Personally, how do you understand intellectual merit in your own research?	"The intellectual merit basically is about the research. It's about what is the research question? And based on that research, what will the answer get us? I always tell people the simple question is who cares? Does anybody care that you're going to do this research? Because if nobody cares, then don't do it, because what's the point? So that's the way I tend to look at intellectual merit." (P06)
Support	Interview Question: How have you been or how could you be further supported in proposing or doing Broader Impacts?	"However. I will say that I in the grants that I've submitted, I do feel like I'm doing this alone, like I'm... I wish there was a little bit more hand-holding, let's say, for the beginning, because this isn't- I don't always feel I have any support from any of my supervisors." (R09)
Analytical Codes		
Weighting	How are BI and IM weighted (in terms of importance, value, necessity, function, and feasibility) in relation to or against each other? Includes statements around how the NSF is perceived to weigh each criterion, or how the participant or their colleagues might weight the criteria.	"And of the two criteria, the Intellectual Merit and Broader Impacts that NSF grades us on when they decide whether to fund a proposal or not, this one probably carries 90% of the weight and Broader Impacts carries about 10% of the weight." (R06)
Concern	Statements referring to concerns, hesitations, or pain points around the merit review criteria in theory, application, or practice. Could refer to BI/IM, or the proposal application and review process generally.	"And I guess my big concern with broader impacts can be that because they can feel frustrating for people, it's well, let's get a quick fix for this. Okay, here are the five things you list that have the most chance of success." (R03)
Institutional Support	Statements referring to institutional support of grant applications and the Merit Review criteria, particularly Broader Impacts. Can include critique of institutional support, recommendations for changes or improvements.	"And of the two criteria, the Intellectual Merit and Broader Impacts that NSF grades us on when they decide whether to fund a proposal or not, this one probably carries 90% of the weight and Broader Impacts carries about 10% of the weight." (R06)

Table 10. Modified Roberts (2009) coding criteria with definitions.

	Roberts (2009)	**Modified Code Definitions**
Criteria for Science		
Infrastructure for Science	Creation of new research methodology, tools, or data sources that will be useful to advance science	Creation of new research methodology, tools, or data sources that will be useful to advance science; **research outputs as broader impacts**.
Training and education	Includes mentoring undergraduates, graduate students, and postdoctoral fellows in the laboratory and teaching classes	Includes mentoring undergraduates, graduate students, and postdoctoral fellows in the laboratory and teaching classes, **as well as training and mentoring early career faculty**
Broadening participation	Recruiting or including under-represented groups in research or in outreach efforts. Includes efforts to attract women to science and to keep them in the academic pipeline for all fields **but excludes funding female students in biology and social sciences.**	Recruiting or including under-represented groups in research or in outreach efforts. Includes efforts to attract women to science and to keep them in the academic pipeline for all fields.
K-12 outreach	Outreach to kindergarten to 12th grade students or teachers helps to get kids excited about science and ensure a pipeline of future scientists.	Outreach to kindergarten to 12th grade students or teachers, helps to get kids excited about science and ensure a pipeline of future scientists. **Includes curriculum development and professional development for K-12 educators and administrators**
Academic collaboration	Research collaborations with other universities in the USA or abroad	Research collaborations with other universities in the USA or abroad, **intra-institutional collaboration across labs or research centers.**
Criteria for society		
Potential societal benefits	Direct claims that the research could help to inform policy, be useful for industry, or lead to some solution to a real-world problem; general statements of improved understanding of a natural or technical process (i.e. climate change or ecosystems) were not included	Direct claims that the research could help to inform policy, be useful for industry, or lead to some solution to a real-world problem; **includes claims that research would benefit society in broad or undefined ways.**
Outreach/ broad dissemination	Dissemination of research results for non-academic audiences in any form (web site, seminars, meetings, newspapers, magazines, etc). Does not include K-12 outreach projects	Dissemination of research results for non-academic audiences in any form (web site, seminars, meetings, **museum exhibitions**, newspapers, magazines, etc). Does not include K-12 outreach projects **or formal school curricula**.
Partnerships with potential users of research results	Includes partnerships with industry, non-profits, community groups, government bodies (local, state, national, or international), and national laboratories	Includes partnerships with industry, non-profits, community groups, government bodies (local, state, national, or international), and national laboratories

To facilitate the coding process, Dedoose QDA software was utilized. Given the exploratory nature of this research and the small sample size, intercoder reliability measures were not implemented. Instead, the research prioritized a detailed and nuanced understanding of each code, with the author serving as the sole coder. To validate the coding accuracy, the participant validation technique of member checking was applied (Carlson, 2010; Candela 2019); select codes and excerpts were shared with three participants (two researchers and one Program Officer) who were familiar with qualitative coding. These participants were selected for member checking both to achieve proportional representation of 15-20% from each group (Carlson, 2010) and because they were the individuals who responded to the request and were available and willing to participate in the validation process. While no significant concerns or comments were noted, any discrepancies identified during participant validation were resolved through discourse and noted in a memo or reflected in the code definition as appropriate.

4. Results

4.1 Survey Results

A total of 922 researchers were sent a link to the survey via email from an official university representative, with the survey's purpose as a part of a dissertation clearly noted. A follow-up reminder sent two weeks following the initial contact. By the end of the survey period, there were 63 complete responses and 60 partial responses, culminating in a total of 123 respondents. Among those providing complete responses, 24 volunteered their contact information for a more in-depth follow-up interview.

The response rate for the survey was 13.3%, which was lower than desired. However, according to Holbrook, Krosnick, & Pfent (2007), response accuracy does not significantly diminish even at rates as low as 5%. Consequently, while the modest

response rate suggests a need for careful interpretation of the results, particularly at general levels, the data collected nevertheless provides a credible insight into individual academic researchers' views on the NSF's Merit Review criteria. A statistical analysis was also conducted to compare the responses of interview participants with those of the broader survey population, revealing no statistically significant differences in the average responses to survey questions among those who participated in an interview and those who did not; Likert question responses can be seen in Tables 6 and 7 of the previous chapter.

One key long-form question in the survey asked researchers who had affirmed participation in a grant with a Broader Impacts component to discuss those Broader Impacts activities, asking: “If yes, what was the project and/or what was the broader impacts goal in your own words? (If multiple projects, please list as many as you are willing to provide).” Responses were then qualitatively coded using deductive thematic analysis (Saldaña, 2014) following a modified coding scheme from Roberts (2009); multiple codes could be applied to the same segment. Data was segmented by each individual project listed by researchers where possible, with multiple codes able to be applied to one project. If individual projects were not differentiated, the researcher’s entire response was taken as one unit for coding.

In total, researcher respondents listed 43 independent research projects that they had executed during grants they had participated in as a researcher, with 129 broader impacts elements. The remaining responses did not differentiate by project and were coded as single units. The percentage of types of Broader Impacts activities performed by researchers is shown in Table 11 as follows:

Table 11. Types of Broader Impacts performed by researchers.

Type	Percent of Examples
Criteria for Science	
Infrastructure for Science	23.26%
Training and Education	17.05%
Broadening Participation	14.73%
K-12 Outreach	5.43%
Academic Collaboration	4.65%
Criteria for Society	
Potential Societal Benefits	13.18%
Outreach/Broad Dissemination	12.4%
Partnerships	9.30%

Infrastructure for Science was the most common example of a Broader Impact provided, including statements like "The broader impacts included improving sustainability through the research output" (S05). Training and Education was the second most common code, with listed examples of Broader Impacts from past projects such as "We trained local undergraduate and graduate students" (S34), and "I oversaw a graduate student" (S51). Broadening Participation was the third most common category of Broader Impacts identified in responses, seen in statements like "To increase diversity within the National Ocean Science Quiz Bowl's Regional Hurricane Bowl" (S60). These findings differ slightly from Roberts (2009) and Woodson and Boutilier (2023), who both used the same coding scheme and found that Training and Education was the most common category.

The code co-occurrence matrix (Figure 9) shows cases where multiple codes were applied to the same excerpt together; the color key provided classifies occurances by frequency. Broadening Participation and Training and Education, in addition to being the second and third most common types, also had the strongest co-occurrence among types of Broader Impacts coded for, as seen in Figure 9. "I used to simply describe how a project would address the criterion, such as by training diverse students." (S47) While Pokorny et al. (2018) describe several issues with the co-occurrence matrices as a means of analyzing qualitative data, including missing

relationships between nearby codes, by focusing on Broader Impacts on a project-by-project basis and coding each project as an entire unit with multiple codes applied as relevant, code co-occurrence can be considered to sufficiently demonstrate patterns within Broader Impacts-type relationships within a project.

	Academic Collaboration	Broadening Participation	Infrastructure for Science	K-12 Outreach	Outreach/Broad Dissemination	Partnerships	Potential Societal Benefits	Training and Education
Academic Collaboration		2	3		1	1		2
Broadening Participation	2		4	1	2	3	2	12
Infrastructure for Science	3	4			1	2		6
K-12 Outreach		1			2	1		2
Outreach/Broad Dissemination	1	2	1	2		4	5	5
Partnerships	1	3	2	1	4		3	3
Potential Societal Benefits		2	5			3		5
Training and Education	2	12	5	2	6	3	5	

Number of Instances	Key
0	
1	
2	
3	
4	
5	
6	
7+	

Figure 9. Code co-occurrence matrix for Broader Impacts examples collected from researcher survey responses.

The strong co-occurrence between the Training and Education and Broadening Participation categories shown in Figure 9 can be seen in examples such as this respondent, who stated the Broader Impacts goal of one project was to "Expose students from groups historically excluded in the science to training and research opportunities. Improve education and recruitment of underrepresented

people into the sciences." (S53). This and similar statements highlight the connection between the Broadening Participation and Training and Education categories.

However, as the survey responses, and resulting codes, were reliant on researchers' recollections of their projects and were often answered in a short amount of time, the descriptions provided of past Broader impacts goals were necessarily superficial to some degree. As the depth of responses could not reasonably be expected to match the complete proposals analyzed by previous studies, this represents a notable limitation.

These results are therefore offered with caution, as the depth and detail of the responses are important for establishing a nuanced understanding of researcher's interpretation of broader impacts, and may account for some of the observed difference in results compared to the literature (Roberts, 2009; Woodson and Boutilier, 2023). As Woodson and Boutilier (2023) also note, though, those projects proposed in the initial grant do not always closely reflect those actually conducted; in this way researcher's oral reports may ultimately more closely reflect the reality of Broader Impacts in practice.

Another challenge in coding the Broader Impacts examples provided by the survey respondents was the use of vague or unclear language, often aspirational, to describe Broader Impacts, a challenge also described by Woodson and Boutilier (2023). The prevalence of vague or aspirational language was somewhat unexpected, as survey respondents were asked to recount Broader Impacts goals from previously complete projects, where more concrete descriptions of actual outcomes and impacts could be reasonably expected. Recognizing the challenges posed by the brevity of the survey format, follow-up interviews with researchers were then organized to probe more deeply into researcher's experiences with and perceptions of Broader Impacts as an aspect of their research.

4.2 Interview Results

Acknowledging the limitations of the survey responses, more detailed follow-up interviews with researchers were then conducted to probe more deeply into researchers' experiences with and perceptions of Broader Impacts (See Appendix B for interview protocol). In total, 24 researchers provided their contact information to be interviewed, and 11 researchers ultimately participated in follow-up interviews. Additionally, while both current and former Program Officers were contacted about participating in this research project, however only former Program Officers opted to be interviewed. In total, seven former NSF Program Officers were interviewed, following a similar interview protocol as the researcher group with language and select additional questions tailored to their roles as Program Officers (see Appendix C).

4.2.1 Broadening Participation Equals Broader Impacts

A recurring theme in interviews with former Program Officers is the reduction of the scope of Broader Impacts to simply include efforts aimed at broadening participation of underrepresented groups in STEM. While these efforts are crucial, among former Program Officers interviewed in particular, there were concerns that researchers were being overly narrow or reductionist in their focus on Broadening Participation, with little actual benefit to themselves or society. Here, a former Program Officer from an Engineering-related discipline emphasized the frequency with which they encounter this conflation of Broader Impacts and Broadening Participation, stating:

> P06: “And the first thing I tell people is this is not just bringing Black people into STEM, okay? This is there's more to broader impacts than increasing and broadening participation by underrepresented groups in STEM. There's a lot you can do. I think it's one of those things that they attribute broader impacts and broadening participation as the same thing. And one is a subset of the other and a very small subset of the other. So yeah, it needs to change. [...] Let me say it this way. They think that all they have to do is broaden participation in order to achieve that goal.”

Another former Program Officer echoed similar experiences with Principal Investigators interpreting Broadening Participation as Broader Impacts in their own Physical Sciences-related field, noting:

> P02: “I mean, every proposal in our field will make some kind of broadening participation argument with, you know, sort of varying levels of plausibility. [...] I realized that there are a lot of people out there who have this misconception that broader impacts equals broadening participation. So instead of saying like, our cleaner water is going to help society, they're like, I'm going to recruit some undergraduate researchers from an HBCU. Like that's the only broader impact that they can think of.”

The above quote emphasizes Program Officers’ observations regarding applicants’ interpretations of the Broader Impacts criterion, particularly in equating Broader Impacts exclusively with broadening participation. While broadening participation is part of the NSF’s guidance on Broader Impacts (National Science Foundation 2021), it is merely one aspect of the larger mandate described by the NSF. The narrow focus of applicants on broadening participation could be attributed to a lack of awareness or understanding of the full range of Broader Impacts activities, or potentially an implicit assumption that broadening participation and assorted diversity initiatives are the most straightforward way to address the Broader Impacts requirement within their projects.

In fact, as both of the above quotes highlight, the mentioned examples are not just broadening participation, but doing so in a very specific fashion - broadening participation in higher education and academia. Proposing to train the next

generation of researchers offers researchers the chance to meet a Broader Impacts requirement while working within a paradigm they are already familiar with. This is perhaps an unsurprising trend given the hesitance and challenges researchers still note when it comes to 'doing' Broader Impacts, with 42% of survey respondents noting they found actually including Broader Impacts in their research to be "extremely, very, or moderately challenging".

This conflation of Broader Impacts and broadening participation in research and academia also echoes the trend evidenced in the code co-occurrence matrix (Figure 1), which showed a strong correlation between the codes for Training and Education and Broadening Participation within Broader Impacts activities, encompassing efforts aimed at training underrepresented minorities in research projects. Overall, likely due to their unique position overseeing not just their own work but the work of a whole field, Program Officers were particularly aware of a larger trend whereby applicants conflated Broader Impacts with the concept of Broadening Participation, and often specifically broadening participation in higher education and academia.

4.2.2 K-12 Outreach Equals Broader Impacts

Interpreting Broader Impacts as broadening participation was not the only observed theme narrowing the scope of Broader Impacts activities, with researchers and former Program Officers also highlighting a trend towards the conflation of Broader Impacts with outreach, particularly in the K-12 space. As one former Program Officer shared in frustration:

> P07: "The big thing that I think is widely misunderstood, and this is probably especially in technical engineering proposals is that broader impacts equals outreach equals… That it's that you have to go to a middle school and talk about your really complicated theoretical engineering thing to have broader impacts. PIs have a hard time really articulating the benefit to society and not just like the benefit to… They think underrepresented groups basically."

These frustrations with Broader Impacts being conflated with K-12 outreach were also shared by researchers interviewed, with one commenting:

> R05: “All I can say is that there's lots of projects that are funded in which the researchers will simply give 1 or 2 presentations in a high school and call it broader impacts.”
>
> DW: And you would not classify that-?
>
> R05: “I mean, is it a broader impact? I mean, maybe there's like some... I don't know if you inspired one kid, like, maybe there's a broader impact? But I don't think the full potential of a project is unlocked unless you think about the broader impacts as integral to what the project actually is, okay?”

Both of the above quotes reflect a pattern seen in multiple interviews, where frustration is aimed not so much at the idea of doing outreach with middle schoolers, but the lack of integration of Broader Impacts into the research as a whole. Four of the seven interviewed former Program Officers and five of 11 researchers made similar statements concerning K-12 outreach, captured under the code “K-12 Outreach”. While the NSF recognizes both K-12 Outreach and Broadening Participation as possible types of Broader Impacts activities, as shown in Table 1, participants’ frustration does not seem to be with the base inclusion of these projects. Rather, the issue lies more with the perception of these efforts as add-ons, rather than opportunities for intentional, tailored diversity initiatives that are thoughtfully integrated into the research - unlocking the ‘full potential of a project’, as quoted above. The question therefore, is how to shift from viewing these activities as mere box-ticking exercises to embedding them as fundamental, impactful components of the research agenda?

4.2.3 “Plug In” Broader Impacts and BI Centers

One emergent strategy for creating and conducting Broader Impacts activities is the establishment of centralized Broader Impacts centers and institutions within

research universities. These entities are often aimed at providing a centralized network of partnerships and resources for Broader Impacts activities, ranging from facilitating collaborations to offering pre-packaged "plug in" or "plug-and-play" activities that research can include without direct involvement, or at least minimal involvement, in creation and execution.

This approach elicited mixed reactions from both Researchers and former Program Officers interviewed, and was often a topic raised by the interviewees themselves. Some were against institutionalized Broader Impacts, citing concerns about reduced faculty involvement:

> R07: "The models that I've seen silo it. They take 5% away and say, now we're going to go talk to third grade classrooms with this 5%, or we're going to do a much better job of talking to third grade classrooms with this 5%, because we'll have an evaluator and all of this sort of stuff. Right. Mhm. It allows the faculty member to not have to think about broader impacts at all. So I'm wholly against it. So I've been part, we have gone through these conversations at least three times at scale, you know, enterprise sort of scale conversations about these broader impact institutes sort of things. And I've put forth models that involve deep engagement by the faculty that are contributing. But everybody's like, no, the point is to get them out of the mix. I was like, well, then the point is, is that you're missing the point."

Another researcher echoed concerns that such efforts would lead to a disconnect between the Broader Impact project and the research itself, specifically citing the example of the University of Colorado Boulder's Broader Impacts Network (BIN), which is described as "CU Boulder's central campus source connecting the research and creative works community with the partners, tools, and resources needed to successfully develop and implement socially significant broader impacts." (University of Colorado Boulder, 2018). This researcher highlighted their own personal concerns around the evolution of Broader Impacts in research generally, and the use of centers and networks to conduct that work specifically, saying:\

> R03: “I think broader impacts still need to connect to the specific research project that you're proposing. And I think the way to highlight that logic would be to sort of counterpose the inverse - what if I had an Intellectual Merit bank, “these are the things that you do in a project about”, and you would say, well, that's silly. It needs to be more specific to your project. And I would just say that's the same thing for Broader Impacts. Now, I'm not trying to besmirch the BINS network here, in some sense what they're doing is a sort of interesting aggregation of broader impacts and how they're articulated and I think that can be valuable. The concern about [The BINS Network] is that that’s a purely technocratic, functional solution that will give you good Broader Impacts in the sense that they're statistically likely to be the kind that get funded. But I don't know if it gets at the issue of treating Broader Impacts like an equal part of a proposal to the Intellectual Merit.”

This researcher also raised a connection between Broader Impacts and the question of funding, pointing out a perceived relationship between ‘good’ Broader Impacts as those “statistically likely to be the kind that get funded”. This question of funding was also raised by other researchers with regard to the role and purpose of Broader Impacts in the proposal and award process. Specifically, another researcher referenced mixed messaging as to the role of funding and greater purpose of including Broader Impacts in research:

> R08: “And, you know, it depends on what your goals are, right? So if your goals are running an effective science communication shop, and you see 5% of the money of these grants being wasted on really bad communication or just not even being used at all for that end, then it makes a ton of sense. But if the goal of BI is to help produce an opportunity for faculty to reflect on their work and what it means to society, well, you sure as hell [are] not going to do that by scraping 5% off the top and running away with it!”

While the majority of comments from researchers and former Program Officers alike indicated hesitance around the creation of Broader Impacts centers and institutes as seen in the examples above, select others were more supportive of the idea, considering it as more of a model for facilitating partnerships and overcoming a lack of experience or skill in ‘doing’ Broader Impacts. One early career professor indicated that the NSF could best support their research efforts by “having repositories where people who are not gifted at Broader Impacts can plug in to pre-

existing programs instead of reinventing the wheel all the time." (R02). That lack of experience, of being "not gifted at Broader Impacts" was a concern among multiple researchers, who stated that greater support, often in the form of official networks or partnerships, would be beneficial as seen in the quote above. Similarly, one former Program Officer, who had previously proposed a "BI Center" at a past institution, was strongly in favor of such efforts, saying:

> P06: "I talk to new faculty here every year on broader impacts and how to achieve it. You know it's not part of their thought processes. Their thought processes are honestly wrapped up in the research. Right? Which is fine. I mean, you know, that's absolutely fine. So what I try to do is tell them, you know, you can find a partner to help you with the broader impacts. [...] You don't have to do all that on your own. And it can achieve greater benefits if you let somebody who maybe knows what they're talking about actually do it instead of you, who knows nothing, right?"

This final sentence highlights a key aspect of both perspectives: navigating a lack of training and skill at 'doing' broader impacts. Should training researchers to understand, value, and perform broader impacts take precedence over the actual impact achieved? Beyond merely a rhetorical question with no one answer, this forces the point of *why* Broader Impacts matters so much as to be required in every NSF grant application - is the goal to foster a deeper, innovative understanding of societal impact among researchers, or to ensure that such impacts are realized, regardless of who executes them? That the establishment of Broader Impacts centers has provoked the mixed reactions seen in interviews with both researchers and former Program Officers highlights an ongoing divergence in how researchers and institutions envision the role and execution of Broader Impacts activities and initiatives that must be reckoned with.

5. Limitations and Generalizability

The overall work and generalizability of the findings from this study are subject to several limitations. Most notable is the issue of respondent bias, particularly social desirability bias, wherein both active researchers and former Program Officers may have been inclined to present themselves in a more favorable light (Nederhof, 1985; Larson, 2019). This concern arises from a reasonable potential reluctance to express negative views about the Broader Impacts criterion with regard to their professional roles, as well as the possible professional consequences of articulating controversial opinions, even when those opinions are anonymized. Such biases may lead to data that skews toward a more positive depiction of the merit review process than is actually the case.

Furthermore, self-selection bias represents another challenge, where individuals holding strong opinions about the NSF criteria can be considered more likely to have participated in the survey or interviews (Bethlehem, 2010). This potential skew towards positively inclined opinions about Broader Impacts could result in an overrepresentation of more extreme or emphatic viewpoints, which may not accurately reflect the broader range of attitudes and experiences within the academic community. Other research projects on community perceptions of Broader Impacts report similar concerns (Watts et al., 2015; Hallinen, 2014).

The generalizability of the study is additionally constrained by its relatively narrow and homogeneous sample size. The research encompassed interviews with only 11 researchers, all sourced from a single research university. This single-site study design inherently limits the applicability of the researcher-specific findings across different academic or institutional contexts. Furthermore, the survey design asked participants to recall Broader Impacts activities potentially across multiple projects and years; given the necessarily brief responses likely provided under time constraints, they are not expected to reflect the same depth as a full project proposal.

As such, analysis of Broader Impacts offered in the survey responses should be taken with caution and such differences may explain deviations from existing literature (Roberts, 2009; Woodson and Boutilier, 2023).

Given these factors, the insights derived from this exploratory research, while valuable, must be approached as a starting point for further research rather than a summative description of the whole research community's attitudes towards the Merit Review criteria. Pending further research that encompasses a more diverse and extensive sample that could potentially support the preliminary conclusions drawn here, these results should be taken as an initial exploration of current perceptions of and experiences with the Merit Review criteria.

6. Discussion and Conclusions

The National Science Foundation set out to offer a flexible, unconstrained suite of approaches to the Broader Impacts criterion, attempting to balance guidance with creativity in order to maximize potential societal benefit with every grant awarded - a process that has historically been met with mixed enthusiasm and skepticism. Yet, the NSF's top-down institutional perspective, valuing flexibility and encouraging risk and innovation, repeatedly ran headlong into the research community's concerns about vague guidelines, and reasonable hesitation to endanger much-needed funding with unclear expectations around Broader Impacts (MacFadden 2019).

The identification of every type of Broader Impacts activity in the survey responses at least once (Table 3) suggests that the NSF's flexible and adaptable framework for the Broader Impacts criterion is achieving its intended effect, at least in part. This breadth of Broader Impact activities serves as a promising indication that researchers are engaging with a wide array of approaches and initiatives, in turn validating the NSF's strategic emphasis on the flexibility of the Broader Impacts

criterion and associated principles.

However, the distribution of those types was uneven, with Infrastructure for Science, Broadening Participation, and Training and Education being the most commonly identified types of Broader Impacts listed by surveyed researchers. A strong co-occurence pattern between Broadening Participation and Training and Education type activities echoes larger themes observed in interviews with researchers and Program Officers, particularly in the academia-specific area of training new researchers. This approach reflects a favorable paradigm for grant applicants where Broader Impacts can occur comfortably in the lab or field in a way that directly contributes to the research output even as it fulfills a Broader Impacts requirement.

Program Officers interviewed repeatedly challenged the conflation of Broader Impacts as Broadening Participation in particular, and both researchers and Program Officers alike expressed frustration with colleagues and applicants who approach K-12 Outreach as another 'catch-all' peripheral solution for Broader Impacts activities. Much of the frustration ultimately focused not on the type of activity, but on the lack of integration of those projects into the research objectives, a frustration that was more common with Broadening Participation and K-12 Outreach than other types of Broader Impacts activities listed. Both groups emphasized the need for a more expansive understanding of, and practical application of Broader Impacts activities that fully leverages the capacity of Broader Impacts to foster societal good.

Program Officers' perspectives on the distribution of Broader Impacts activities and the tendency to focus exclusively on Broadening Participation, K-12 Outreach, and intra-academia training also raised a key difference between the two groups interviewed in this work. Specifically, the difference between the larger-scale perspective required of Program Officers and the necessarily smaller-scale perspective of the average individual researcher. Charged with overseeing all

applications within their divisions, Program Officers must acquire a broader view of research and its progress; several referred to this as a 'birds-eye view' in interviews. That larger perspective may account for some of the frustrated responses to Broadening Participation as Broader Impacts, and also highlights the critical role Program Officers play in bridging individual researchers' practical needs with the NSF's larger institutional expectations (Epstein 2016).

A larger push to involve more researchers in the review process - as rotational Program Officers and reviewers, and by integrating grant writing earlier into graduate education - may benefit the overall arc of research impact over time. Such efforts would facilitate earlier exposure to a greater variety and breath of broader impacts possibilities, and also normalize broader impacts as an integral part of the research process.

At the time of writing, additional Broader Impacts resources and tools abound but most focus on how to put a broader impacts project together on top of an existing Intellectual Merit or research objective. Fewer options are available for constructing broader impacts from the ground up as an integral part of the research process *alongside* and *equal to* Intellectual Merit. Fecher and Kobsda's Research Impact Canvas (2019) is one example of tool that could be used to such effect, and other current options, such as ARIS's Broader Impacts Toolkit (2023c) can be repurposed to similar ends.

However, these tools are only effective if implemented from the outset, and that itself is contingent on researchers and the larger academic community initially valuing and incentivizing the outcomes - impactful, intentional and successful Broader Impacts - that those tools are designed to achieve. Despite earlier hurdles, broader adoption of Broader Impacts seems to demonstrate the gradual success of the National Science Foundation's approach. Ultimately, 'unlocking the full potential' of Broader Impacts, to quote one researcher, will require an even greater shift

amongst scholars and researchers that considers more than just bringing underrepresented minorities into the lab, and instead centers research impact and societal benefit alongside the research objectives from the very first moment.

7. References

Bethlehem, J. (2010). Selection bias in web surveys. *International statistical review*, *78*(2), 161-188.

Candela, A. G. (2019). Exploring the function of member checking. *The qualitative report*, *24*(3), 619-628.

Carlson, J. A. (2010). Avoiding traps in member checking. *Qualitative Report*, *15*(5), 1102-1113.

Center for Advancing Research in Society. (2023a). *What Are Broader Impacts?* Advancing Research in Society. https://researchinsociety.org/broader-impacts/

Center for Advancing Research in Society. (2023b). *Evolution of Broader Impacts*. ARIS. https://doi.org/10.32469/10355/95863

Center for Advancing Research in Society. (2023c). ARIS Broader Impacts Toolkit. Advancing Research in Society. https://aris.marine.rutgers.edu/

Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research*, 50(2), 708-739.

Epstein, J. (2016). Reflections of an NSF Program Officer. *IEEE Security & Privacy*, *14*(2), 3-6.

Fecher, B., & Kobsda, C. (2019). *Meet the Research Impact Canvas: A Structured Guide for Planning Your Science Communication Activities*. https://doi.org/10.5281/zenodo.3368480

Gusterson, H. (1997). Studying up revisited. *PoLAR, 20*, 114.

Hallinen, J. R. (2014). The many quiet tensions: Perceptions of the broader impacts criterion held by NSF career award holders at very high research institutions of higher education.

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, *61*(5), 534-547.

MacFadden, B. J. (2019). *Broader impacts of science on society*. Cambridge University Press.

Mervis, J. 2007. Pilot NSF Program flies into stiff community headwinds. *Science*. 318:1365-1366.

MERVIS, J. (2011). Beyond The Data. Science, 334(6053), 169–171. http://www.jstor.org/stable/23059239

MIT School of Science. (2024, January). *NSF Broader Impacts statement guidance*. MIT School of Science. https://science.mit.edu/resource/nsf-broader-impacts-statement-guidance/

Nadkarni, N. M., & Stasch, A. E. (2013). How broad are our broader impacts? An analysis of the National Science Foundation's Ecosystem Studies Program and the broader Impacts requirement. *Frontiers in Ecology and the Environment*, *11*(1), 13-19.

Namey, E., Guest, G., Thairu, L., Johnson, L., Guest, G., & MacQueen, K. (Eds.) (2007). Handbook for team-based qualitative research. Lanham, MD: Altamira Press.

National Science Foundation (1997). *Important Notice No. 121* (in121). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in121.txt

National Science Foundation (1999). *Important Notice No. 125* (in125). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in125.pdf

National Science Foundation (2002). *Important Notice No. 127* (in127). Washington, DC. Accessed 15 August 2024 at https://www.nsf.gov/pubs/issuances/in127.pdf

National Science Foundation (2007a). (Dear Colleague Letter) Broader Impacts Review Criterion (NSF07046). Washington, DC. Accessed 10 May 2024 at https://www.nsf.gov/pubs/2007/nsf07046/nsf07046.jsp

National Science Foundation (2007b). *Proposal and Award Policies and Procedures Guide* (NSF 07-140, GPG07140). Washington, DC. Accessed 13 June 2024 at https://www.nsf.gov/pubs/policydocs/papp/gpg07140.pdf

National Science Foundation (2013). *Proposal and Award Policies and Procedures Guide* (NSF 13-1, NSF13001). Washington, DC. Accessed 13 June 2024 at https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/nsf13_1.pdf

National Science Foundation (2020). *Proposal and Award Policies and Procedures Guide* (NSF 20-1, NSF20001). Washington, DC. Accessed 13 June 2024 at https://www.nsf.gov/pubs/policydocs/pappg20_1/nsf20_1.pdf

National Science Foundation (2023). *Proposal and Award Policies and Procedures Guide* (NSF 24-1, NSF24001). Washington, DC. Accessed 22 June 2024 at https://nsf-gov-resources.nsf.gov/files/nsf24_1.pdf

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology*, *15*(3), 263-280.

Notre Dame. (2024, July 16). *Center for Broader Impacts*. Center for Broader Impacts. https://cbi.nd.edu

NSB-NSF Staff Merit Review Task Force. (1997, March). Merit Review Final Recommendations. https://www.nsf.gov/pubs/1997/nsbmr975/nsbmr975.htm

University of Colorado Boulder. (2018, April 18). *Broader Impacts Network*. Research & Innovation Office. https://www.colorado.edu/researchinnovation/research-development/other-resources/broader-impacts-network

University of Wisconsin-Madison. (2024). *Broader Impacts Homepage*. University of Wisconsin-Madison. Retrieved August 19, 2024, from https://broaderimpacts.wisc.edu/

Pokorny, J. J., Norman, A., Zanesco, A. P., Bauer-Wu, S., Sahdra, B. K., & Saron, C. D. (2018). Network analysis for the visualization and analysis of qualitative data. Psychological methods, 23(1), 169.

Princeton. (2024, July). *Supporting STEM Broader Impacts*. Council on Science and Technology. https://cst.princeton.edu/initiatives/broader-impacts-stem-outreach

Risien, J., & Storksdieck, M. (2018). Unveiling impact identities: A path for connecting science and society. *Integrative and Comparative Biology*, *58*(1), 58-66.

Rothenberg, M. (2010). Making judgments about grant proposals: A brief history of the merit review criteria at the National Science Foundation. *Technology & Innovation*, *12*(3), 189-195.

Saldaña, J. (2014). Coding and analysis strategies.

Virginia Tech. (2024, July). *Broader Impacts Partnerships*. Center for Educational Networks and Impacts. https://ceni.icat.vt.edu/content/ceni_icat_vt_edu/en/partnership-process.html

Watts, S. M., George, M. D., & Levey, D. J. (2015). Achieving broader impacts in the national science foundation, division of environmental biology. BioScience, 65(4), 397-407.

Woodson, T., & Boutilier, S. (2023). From intent to impact—The decline of broader impacts throughout an NSF project life cycle. *Research Evaluation*, *32*(2), 348-355.

CHAPTER 4

BROADER IMPACTS, OFFICIALLY

Broader Impacts, Officially: The Proceedings of the 2023-2024 NSB-NSF Task Force On Merit Review

1. Introduction

The National Science Foundation (NSF) has relied on the current Merit Review Criteria for nearly 30 years at the time of writing, with Intellectual Merit and Broader Impacts serving as the dual pillars guiding funding review and awards. While previous chapters in this dissertation, building on an extensive body of literature, have examined how individual researchers and Program Officers interpret and navigate these criteria—particularly the Broader Impacts requirement—a critical gap remains in understanding how the NSF as an institution, along with its governing board the National Science Board (NSB), interprets and potentially revises these guidelines. This paper shifts focus from individual experiences with the Merit Review Criteria to considering the top-down, institutional perspective of the National Science Foundation. In doing so, the paper aims to address this gap by analyzing open meeting notes from the ongoing 2023-2024 NSB-NSF Task Force on Merit Review, providing a rare glimpse into the high-level deliberations that could shape the future of research funding in the United States.

Previous studies have largely focused on the individual perspectives of grant applicants and Program Officers, with a particular focus towards the variety of activities proposed (Roberts, 2009) and individuals scholars' motivation (Wiley, 2014; Nagy, 2016) and providing support and resources for scholars looking to write fundable and convincing Broader Impacts Statements (Skrip, 2015, Heath et al., 2014; Cotos, 2019). Such papers also often highlight the challenges researchers face

in meeting the Broader Impacts criterion. These challenges include ambiguity in guidelines (Reilly, 2021), inconsistent application of the criteria (Kamenetzky, 2013), and the perceived burden on researchers to demonstrate societal relevance alongside scientific excellence (Hallinen, 2014). Such critiques are important, as they reflect the day-to-day realities of the academic community that the NSF serves.

However, they also overlook the broader institutional perspective that the NSF must maintain. Unlike individual researchers, who focus on specific projects, the NSF is responsible for the impact of its funded research not just on a project-by-project basis, but in the aggregate as well. In funding projects with an eye not just to the application at hand but also the current state of the various scientific disciplines and their long term direction, NSF staff must ensure that awards and decisions not only advance scientific knowledge but also benefit society as a whole. All the while, the NSF must remain accountable to government stakeholders (Frodeman et al., 2013; Rothenberg, 2010), balancing the need for scientific innovation with the responsible use of public funds.

The 2023-2024 Task Force on Merit Review, known by its acronym MRX and referred to here as alternately the Task Force or the MRX commission, presents a unique opportunity to study how the NSF, through its institutional lens, approaches the ongoing evolution of these criteria (National Science Board, 2023). Task forces such as this one are typically convened to address specific challenges or emerging needs within the NSF's operational framework, and as such offer insights into how the organization interprets its mission and adapts to external pressures, including shifts in societal expectations, policy directives, and the broader landscape of scientific research. Notably, the current task force benefits from a more transparent and accessible process than its predecessors, with meeting notes and recordings made publicly available in unprecedented detail.

Given the valuable access to the internal workings of the NSF and the ongoing nature of the Task Force proceedings at the time of writing, a novel transcription methodology utilizing a Large Language Model (LLM) was employed to efficiently process and clean large volumes of raw text data as videos were released. This approach not only facilitated timely analysis but also presents a valuable method for rapidly parsing publicly available government proceedings as they unfold, which is increasingly important in an era of growing digital data access and availability

The key research question underlying this project was: How and why do stakeholders engaging with the NSF - either as researchers, Program Officers, or institutional officials - construct and value research impact, particularly as a function of the NSF's Broader Impact criteria? Relevant to the NSB-NSF Task Force on Merit Review particularly is the subquestion: How do NSF Program Officers and officials understand and facilitate impact through their role within the NSF application review process?

Ultimately, this paper situates the 2023-2024 Task Force, which is ongoing at the time of writing, within the context of previous efforts to refine the Merit Review criteria, using LLM-cleaned transcripts of open MRX meetings. These include the pivotal 1997 task force that introduced the Intellectual Merit and Broader Impacts criteria, and the 2010-2011 task force that sought to clarify and improve the understanding of, and application of, Broader Impacts particularly. By comparing the current discussions with those from earlier task forces, this analysis hopes to illuminate how the NSF's institutional priorities and external obligations continue to shape its approach to the Merit Review criteria. This paper also seeks to offer a novel angle to discussions of the Broader Impact criterion by considering the NSF's top-down perspective in conjunction with the individual proposal-focused analyses that have been more commonly seen (Nagy, 2016; Roberts, 2009; Hallinen, 2014; Gould et al, 2019; Cotos, 2019).

The role of the NSF in the research process is challenging to parse but critical in understanding the complex network of needs and tensions facing both the individual researcher and the Foundation as an organization and major science funding administrator. As NSF historian Marc Rothenberg concludes in his discussion of the evolution of NSF proposal criteria:

> "There have been four constants in this discussion. Far more worthy proposals are received than are fundable, requiring some criteria to be developed to distinguish among meritorious proposals. Program expansion and external issues have forced the NSF to consider criteria other than simply good science. The peer review community does not necessarily apply the criteria in the manner the NSF wishes. And there are always new programs and new issues forcing the NSF and the NSB to rethink these criteria." (2010)

These needs and tensions are not new, but constantly shifting and evolving with the research landscape, and understanding them now at this moment in 2024 may help us understand the evolution of science funding and policy in the next decade.

2. Historical Context

2.1 Evolution of the Merit Review Criteria

Founded in 1950, the National Science Foundation circulated its first call for proposals in 1951. This call included the first look at the organizations' intended review criteria, which was initially announced as a single criterion, "the scientific merit of the proposed research, including the competence of the investigator". However, in the first year of grant awards, the single criterion was expanded for practical reasons (Rothenberg, 2010) as the first round of proposals flooded in. The final four criteria were, roughly: 1) the intrinsic merit of the research, 2) researcher performance competence, 3) the utility or relevance of the research, and 4) the effect on the infrastructure of science and engineering.

Due to a decade of growth and budget expansion, as well as unequal funding of larger research centers and institutions (Rothenberg, 2010), in 1967 the NSB presented the “Criteria for the Support of Research”, which offered revised criteria that were categorized by the type of institution receiving funding. These criteria mixed technical and non-technical merit, including “educational value”, “potential applications”, and “tangible social benefits” of the work (p. 217-219). At the time of publication, they stated that these criteria constituted "no major departure from current practice" and were instead a "clarification and reaffirmation" of NSF philosophy since its founding, potentially indicating the concerns encompassed by what we now refer to as 'Broader Impacts' likely existed prior to their formal codification into the revised 1967 criteria.

The year 1975 marked a pivotal moment for the National Science Foundation (NSF). Having spent its first 25 years developing a peer review system that operated with minimal controversy, the NSF found itself under intensified public scrutiny in the politically charged atmosphere of the 1970s, shaped by events such as the Vietnam War and the Watergate scandal (Mazuzan, 1992). No longer a small government agency primarily engaged with the scientific research community, the NSF's expanding purview into social sciences, behavioral research, and educational topics exposed it to criticism from groups that had previously ignored its activities. This can be seen most clearly in the controversy surrounding the partly-NSF funded social studies curriculum “Man: A Course of Study”, or MACOS, which put the NSF’s processes directly in the path of Congressional oversight (Mazuzan 1992, Rothenberg 2010).

MACOS, a novel social studies curriculum developed in the 1960s, was a program intended for fifth and sixth grade students that aimed to explore the question “What is human about human beings” using a mixture of scientific and anthropological content, including the lifecycle of salmon and the daily life of Netsilik

Eskimos. The curriculum drew the ire of congressional politicians and constituents who saw federally funded pre-college school curricula efforts as intrusive and diminishing the parental role (Dow, 2013). Congressman John B. Conlan of Arizona - who was previously of the faculty at Arizona State University, the very institution where this dissertation is being written - spearheaded the initial pushback, heading legislative amendments to the annual NSF budget that threatened to defund the entire organization (Mazuzan, 1992). Although these amendments ultimately stalled in Congress, the controversy escalated beyond MACOS, jeopardizing the NSF's entire suite of educational programs as Conlan broadened his attack to encompass all NSF involvement in science education (Dow, 2013), and eventually undermining the credibility of the entire foundation in the process.

Following 1975, the NSF faced ongoing pressure concerning accountability and evaluation, particularly under the Reagan Administration, which succeeded in significantly reducing the scope and budget of the Education Directorate (Dow 2013; Mazuzan, 1992). Having succeeded with limiting educational activities at the NSF, governmental and public attention refocused particularly heavily on the peer review system. The so-called "Conlan Wars," as they were known internally at the NSF (Dow, 2013), led to a high rate of turnover at the upper levels of NSF administration between 1974 and 1989. Coupled with increased public scrutiny, these challenges compelled division directors and successive NSF directors to continually justify the agency's existence, often beginning with evidence of the rigor of the peer review process responsible for allocating taxpayer dollars in the form of grants (Rothenberg, 2010). It was this assertion of Congressional oversight that provided much of the impetus for change at the NSF in decades to come.

2.2 The CHIPS act and Government Accountability

The passage of the CHIPS and Science Act (2022), a major piece of US legislation aimed at bolstering domestic semiconductor manufacturing, strengthening supply chain resilience, and enhancing the nation's leadership in scientific research and innovation, was a direct catalyst for the establishment of the MRX Task Force. In addition to allocating substantial funding for the development and production of semiconductors within the United States the CHIPS and Science Act included provisions to expand and support research and development across various scientific fields. Of particular relevance to the NSF is Section 10343 of the Act, titled "Research Ethics," which directs the NSF to incorporate ethical, social, safety, and security considerations into the NSF's merit review process. In the MRX Task Force's statement of goals, it notes that the CHIPS and Science Act "reinforces NSF's strategic focus on both delivering benefits from research and developing diverse domestic STEM talent, [and] increases the importance of the broader impact review criterion" (National Science Board, 2023).

2.3 Merit Review Criteria Task Forces

Much has been written about how individual scientists and various fields and disciplines understand the Merit Review criteria, particularly Broader Impacts. Less ink has been spent considering the expectations, needs, and tensions around the Merit Review Criteria within the NSF itself, in part due to the practical consideration that there is little public access to or evidence of how NSF staff interpret the Merit Review criteria, and particularly the Broader Impacts criterion, from day to day.

However, a sequence of public Task Forces over the years have occasionally opened a door into the inner thinking within the organization. Occurring roughly every ten years since the introduction of Intellectual Merit and Broader Impacts, the

NSF has convened Merit Review focused Task Forces in 1996-1997, and 2010-2011, with a third currently in progress at the time of writing in 2023-2024.

Both previous task forces brought about some degree of change; the 1997 task force initiated the change to Intellectual Merit and Broader Impacts from the four previous criteria, while the 2010 task force focused on providing increased guidelines and resources on broader impacts in response to concerns raised by the general research community. However, the impact of those task forces is less clear. The founding of the National Alliance for Broader Impacts (NABI) in 2014 and its later 2018 report "Current State of Broader Impacts: Advancing Science and Benefitting Society" found that, of seven common issues raised by respondents, the majority of those issues had already been brought to the NSF's attention as far back as the initial 1997 Task Force prior to the implementation of the revised Merit Review Criteria (NABI, 2018). Among those commonly raised issues were the question of how the criteria were to be weighted against each other, as discussed in Paper 1, and inconsistency in both judgment and application as a result of NSF attempts to allow for 'flexible' weighting and application (Bowker, 2020).

2.4 Current NSB-NSF Commission on Merit Review

The NSB-NSF Task Force on Merit Review, also known as the MRX Commission, was established at the December 1-2, 2022, National Science Board meeting to evaluate the efficacy of the NSF's existing Merit Review policy, criteria, and processes. At this time, the MRX Commission was tasked with examining how effectively the current Merit Review Criteria, principles, and elements supported the NSF's mission. This reexamination was motivated by multiple concerns including the balance between Intellectual Merit and Broader Impacts, persistent challenges in measuring research outcomes, and disparities in funding, particularly those affecting underrepresented researchers. Furthermore, the MRX commission charge notes that

"In the past decade, three laws have expressed relevant Congressional intent and interest" (National Science Board, 2023) in the Merit Review process and the selection, funding, and evaluation of awards within the NSF. In particular, the passage of the CHIPS and Science Act highlighted the growing importance of the Broader Impacts criterion as noted above.

3. Methods

This study examines how the National Science Foundation, as represented by its official representatives and institutional officials, currently understands the Merit Review process, particularly the Broader Impacts criterion, and how this process may be reformed as a result of the ongoing NSB-NSF Task Force on Merit Review. These potential reforms could have long-term implications for the implementation of research processes and scientific funding mechanisms on a national scale. To explore this, publicly available recordings of open Task Force meetings were collected, transcribed, and analyzed using qualitative coding methods.

3.1 Data Collection:

Each MRX open meeting recording was identified from two sources: first, the official 'Merit Review Reexamination Commission Resources' page (National Science Board, 2024), and second, the official National Science Board YouTube channel (Appendix D). Of all recordings identified, nine were shown on the official resources page and identified on the YouTube channel; another three videos were identified solely from the YouTube channel. That not all videos were stored on the official resources page is potentially due to administrative error, given the wide variability of titles and upload parameters.

Title formatting was inconsistent, with videos listed variously as “NSB-NSF Commission of/on Merit Review”, “Merit Review Commission”, and “Open MRX”. Absent a centralized official collection of all relevant videos, in searching YouTube a combination of publication on the NSB channel and search terms “MRX”, “NSB-NSF Commission” and “Merit Review” were used to identify a more complete list of all recordings on the official NSB YouTube channel, shown in Table 12.

Table 12. Collected open meeting recordings with assigned research ID.

Meeting ID	**YouTube Title**	**Date**	**Recording Length (h:m:s)**	**Meeting Agenda**
MRX 1	OPEN MRX Meeting	19-Apr-23	1:01:26	Workplan and ARIS listening session
MRX 2	Open MRX Meeting, May 8, 2023	8-May-23	2:22:36	Overview of Merit Review Criteria
MRX 3	Open MRX Meeting	24-May-23	58:00	Broader Impacts
MRX 4	Open MRX Meeting	28-Jun-23	1:27:41	Intellectual Merit
MRX 5	Merit Review Commission	14-Aug-23	1:38:13	Federal Agency Perspectives on Grant Review
MRX 6	NSB-NSF Commission on Merit Review Open Meeting	27-Sep-23	58:30	Additional Review Criteria
MRX 7	NSB-NSF Commission of Merit Review - Open Meeting	25-Oct-23	59:36	Criteria Evaluation
MRX 8	NSB-NSF Commission on Merit Review (MRX) Meeting	28-Nov-23	2:03:40	High-Risk High Reward Funding, Portfolio Management
MRX 9	NSB-NSF Commission on Merit Review Meeting	13-Dec-23	55:56	Broadening Participation and DEIA
MRX 10	NSB-NSF Commission on Merit Review (MRX) Meeting	18-Jan-23	1:53:20	Assessing potential policy recommendations
MRX 11	NSB-NSF Commission on Merit Review (MRX) Meeting	29-Jan-24	1:12:35	Assessing potential implementation and accountability recommendations
MRX Plenary	NSB-NSF Commission on Merit Review - Plenary Open	10-May-24	12:39	Introduction of report and recommendations

In total, 11 open meeting recordings were identified. One additional commission video was also collected as listed on the official resources page. This was

an Open Plenary from the 488th Meeting of the National Science Board in May 2024 where the chair, Steven Willard, gave a report on the progress of the commission. The 12 videos totaled 15 hours and 47 minutes runtime, and had been recorded and posted at various points from April 19th 2023 through May 10th, 2024. Table 12 shows the list of meetings and dates posted on YouTube, as well as the recording length and the official meeting agenda described for each meeting. A link list of all meeting recordings is provided in Appendix D. All commission meetings will be referred to using their assigned Meeting ID in the remainder of this work. Notably, these are not the only times the Task Force met, with a number of closed meetings also occurring in addition to the open meetings posted to YouTube; the number of closed meetings could not be accurately determined from the official National Science Foundation calendar or schedule.

3.2 LLM-assisted Transcription Process

Transcription efforts in this portion of the research were performed following a denaturalistic approach to transcription as discussed by Oliver et al. (2005), that focuses on the meaning of speech more so than its linguistic patterns. A denaturalized transcription approach was selected to prioritize the semantic content of the interviews over the intricacies of speech patterns, ensuring that the transcripts remain clear, readable, and legible for an analysis focused on meaning rather than the nuances of verbal delivery, in line with the research focus on participant's perceptions of the topic. Given this more hands-on and edited approach, a greater degree of transcription cleaning than a naturalized approach typically demands was required. As such, a combination of traditional hand-cleaning and LLM-assisted transcription cleaning methods were applied to each recording's audio.

In order to effectively and efficiently analyze nearly 16 hours worth of commission meetings, a novel method for using a Large Language Model (LLM),

specifically the LLM used by OpenAI's ChatGPT 4.0 (Open AI, 2024), was implemented prior to traditional hand cleaning, with the goal of streamlining bulk formatting and cleaning. The decision to use ChatGPT to do a first-pass at cleaning each transcript was made in part due to the challenges posed when transcribing YouTube files.

YouTube does offer a native 'auto-generated' transcription option for any video containing speech, shown in Figure 10, here boxed in red at the right of the screenshot. This transcription is automatically generated using speech recognition technology, and YouTube officially defines their transcripts as "Unformatted (and untimed) text that's transcribed verbatim from the video" (Google Support 2024a).

Figure 10. Screenshot of YouTube in browser, showing active 'Transcript' option.

The transcript shown in Figure 10 can be copied and pasted, but results in an inefficient and challenging-to-read format, as seen in Table 11, Step 1 below. This format was not conducive to reading or the intended analysis via qualitative coding. Consequently, ChatGPT 4.0 was used to assist in bulk reformatting and transcript cleaning.

First, each video was given an identification as MRX 1-11 or MRX Plenary in sequential order (Table 13), and the native YouTube transcript of each video was copied to a standalone document. In order to account for the token limit on ChatGPT 4.0, which limits the possible length of the message sent and the length of the model's response, each file was divided into sequential 700-900 word 'blocks', with each break between blocks placed at an apparent pause in dialog - either a change of speaker or the assumed end of a sentence. The approximate location of each break was identified based on surrounding text and context clues. Each block was then uploaded into ChatGPT 4.0, and each block submitted was preceded by the following prompt:

> *"The following text after the paragraph break is transcribed from an NSB meeting about the NSF merit review criteria. It needs to be properly punctuated, with proper capitalization. Please add paragraph breaks. Please remove time stamps. Please remove 'uh', 'um', and repeated words such as 'if if', but otherwise please use as much of the original text as possible.*
>
> *The text to be cleaned up is as follows:"*

The prompt was engineered following a denaturalized approach to transcription, hence the removal of stutter words such as 'um' and other repeated words, prioritizing meaning and readability over exact linguistic representation (Oliver et al., 2005). All blocks from one video were transcribed sequentially within the same chat, and a new chat was started for each video for organizational purposes. Initial testing indicated that only the first block submitted to a chat window could include the prompt, and that ChatGPT would typically continue to apply the same prompt to subsequent blocks submitted with no prompt attached for several rounds. However, including the prompt with every block of text reduced the anecdotally observed number of hallucinations, here taken as seemingly correct but non-factual or inaccurate generated responses (Alkaissi and McFarlane, 2023; Siontis, 2024), that occurred.

Table 13. LLM-Assisted Transcription steps, with text examples.

Transcription	**Text Excerpt**
Step 1: YouTube Transcript	5:50 self-employments or other next we then ask participants to share their 5:58 thoughts on the overall strengths and weaknesses of the Vi Criterion 6:04 as you can see in this word cloud participants used words like societal 6:09 benefit Community engagement inclusive required Broad and 6:17 accountability to describe the strengths of the Criterion 6:22 in discussion participants also gave NSF credit for making bi a requirement in 6:29 proposals as it encourages pis and institutions to pay attention to bi in 6:36 funded research next as for weaknesses participants used
Step 2: LLM-Cleaned Transcription	Next, we then ask**ed** participants to share their thoughts on the overall strengths and weaknesses of the **BI** criterion. As you can see in this word cloud, participants used words like "societal benefit," "community engagement," "inclusive," "required," "broad," and "accountability" to describe the strengths of the criterion. In discussion, participants also gave NSF credit for making BI a requirement in proposals, as it encourages PIs and institutions to pay attention to BI in funded research.
Step 3: Final Transcript	**[5:50] WW:** Next**!** We then asked participants to share their thoughts on the overall strengths and weaknesses of the BI criterion. As you can see in this word cloud, participants used words like "societal benefit," "community engagement," "inclusive," "required," "broad," and "accountability" to describe the strengths of the criterion. In discussion, participants also gave NSF credit for making BI a requirement in proposals, as it encourages PIs and institutions to pay attention to BI in funded research.

While LLM hallucinations were only occasionally observed during the LLM-aided transcription process, they did occur with enough frequency to warrant checking every response from ChatGPT carefully before adding it to the complete LLM-cleaned document. Hallucinations occurred most often when the leading sentence of the text block was a question or other direct piece of dialog that the system would take as a question or statement for it to respond to, rather than text to be transcribed.

The following excerpt was the starting portion of text in a block from MRX 8, reading:

> "I'm I'm trying to pick out out of this list that we were given oh yes which you don't have in front of no I don't have that list sorry yeah um you know this is the kind of thing I would think would be a natural to include in the Merit review digest unless I'm missing it or maybe it's there" (MRX 8)

Submitted *without* the prompt, ChatGPT instead generated a hallucinatory response that began:

> "Thank you for bringing that up. Indeed, the issue of how portfolio balance is documented and understood within the NSF, especially in relation to agency priorities, is crucial."

ChatGPT's hallucinated response continued in a similar vein. When re-prompted *with* the prompt and the same block of text, a formatted transcription was then correctly generated as intended, as a portion of dialog between two Task Force members trying to navigate their internal resources repository to access a particular document, shown below with speaker names withheld:

> Speaker 1: "I'm trying to pick out from this list that we were given—oh yes, which you don't have in front of—"
>
> Speaker 2: "No, I don't have that list, sorry. Yeah."
>
> Speaker 1: "You know, this is the kind of thing I would think would be a natural to include in the Merit Review digest, unless I'm missing it or maybe it's there." (MRX 8)

Hallucinations that occurred were almost entirely in this fashion, responding to a supposed question within the text to be transcribed and thus beginning the response as such; no cases of a complete hallucination manifesting part way through a response or responses that mixed both accurate transcription and large-scale hallucination were observed.

An example of an otherwise reasonably correctly generated LLM-cleaned transcript can be seen in Table 13, Step 2. Note that the majority of changes made to the transcript were the addition of capitalization and punctuation, with other minor changes being the correction of “Vi” to “BI” and the change of tense from ‘ask’ to ‘asked’. Underlined and bolded text indicates where a material change was made to the text. Capitalization, punctuation, and formatting changes are not indicated.

After being run through ChatGPT, and initially verified as correctly generated, all LLM-cleaned text blocks were then recombined into a complete document for further cleaning. LLM-cleaned transcripts were then compared to the original video audio to verify accuracy, correct word choice and punctuation, identify speakers, and apply timestamps. This second-pass hand cleaning was also used as another level of verification that no hallucination or otherwise incorrect or skipped data had been missed during generation.

As seen comparing Steps 2 and 3 in Table 13, the majority of changes from the LLM-cleaned version to the final hand-cleaned version were focused on nuances of phrasing - i.e. where the first word ‘next’ was initially taken to be the start of a sentence, in actuality it was a request for the next slide of a presentation, necessitating a change of punctuation. Other changes, such as ‘ask’ to ‘asked’, were verified as correct via the original audio.

3.3 Flexible Qualitative Coding

After meeting audio was collected and the complete transcription process was finalized, a qualitative coding approach was applied to all transcripts commission meeting notes, with the goal of identifying themes concerning the National Science Foundation’s perspective on, concerns with, and possible changes to the Merit Review Criteria. Select codes are shown in Table 14.

Table 14. Selected in vivo codes with definitions and examples.

Code	Definition	Example
Portfolio Balance	Statements referring to balancing the development, management, or composition of a portfolio of awards within the NSF. May focus on specific facets of the portfolio, or on the decision-making process by Program Officers and other staff. May be within a single division or across the institution.	There are multiple points of interaction around any individual award decision, and that issue or question of portfolio balance is there are many individuals that are part of that. (MRX 8)
Flexibility	Statements referring to flexibility in guidelines, application of the criteria, and in research processes more generally. May refer to either Merit Review criteria.	Yes, you want flexibility, and you don't know when the ideas are going to come along and so forth, so I think an openness and an encouragement, with appropriate checks and balances, is really important agency-wide. (MRX 8)
Institutional Support	Statements referring to institutional support of grant applications and the Merit Review criteria, particularly Broader Impacts. Refers to research institutions, institutions of higher education, and similar. Can include critique of institutional support, recommendations for changes or improvements.	With regard to institutional support for BI, this was another emerging theme, with participants expressing varying degrees of support for BI from their institutions. They suggested that institutions of higher education do an inadequate job, if at all, at tracking and documenting the impact of BI, whether at the PI, the departmental, or the institutional level. (MRX 3)

Deterding and Waters' "Flexible Coding of In-depth Interviews: A Twenty-first-century Approach" (2018) informed the qualitative coding approach taken in this portion of the project. While the data in question was not interviews, but rather

recorded open meeting notes, the 'flexible coding' approach implemented in other portions of this research allows for modifications in the coding scheme as more understanding is gained about the data. This would assist in ensuring that the analysis remained relevant and comprehensive even as the commission modified their own definitions of and approach to the Merit Review criteria over the course of the 13-month time span covered by the analyzed video recordings.

Additionally, In Vivo codes were applied to the transcripts as analysis progressed, with examples provided in Table 14 showing In Vivo codes as defined, and text excerpts involving the code. Given the complexity of the text analyzed, simultaneous codes were also applied in cases where multiple codes appeared in one segment. Dedoose QDA software was used to facilitate coding. Intercoder reliability was not performed due to the small sample size and exploratory nature of the research. With the author acting as the only primary coder, a focus on a detailed and nuanced understanding of individual codes was prioritized over generalizability.

4. Results

4.1 LLM-assisted transcription

The LLM-assisted transcription process yielded transcripts suitable for analysis with a reasonably modest amount of effort required, most of which was required primarily in the final hand-cleaning stage. This process anecdotally resulted in cleaner and more accurate transcripts in less time compared to the more traditional methods used for transcribing interview data in other phases of this research.

The bulk of the time and effort was devoted to identifying speakers and adding timestamps for each new speaker, as opposed to the more frequent but less relevant timestamps originally provided by YouTube. Given this, a smaller amount of time was ultimately required for editing the spoken content itself. It is important to

note that the efficiency of this transcription process can be partly attributed to the nature of the speech captured during the MRX commission meetings, which was characterized by structured presentations and prepared comments that often transcribed reasonably cleanly.

In contrast, when off-the-cuff questions and responses from committee members were offered, these often included run-on sentences, pauses, and other verbal artifacts, and were more challenging to transcribe and required more time and effort during cleaning. Traditional qualitative research interviews—such as those transcribed in other phases of this project—often more closely resembled that second kind of 'off-the-cuff', unprepared speech. Such material involving participants less familiar with public speaking or addressing topics on which they do not have concise, prepared statements, can lead to inherent differences in the quality and complexity of the speech requiring transcription, and may not have benefitted so clearly from the application of LLM-assisted transcription cleaning methods.

4.2 MRX Open Meetings

In discussing the proposed workplan of the MRX commission during MRX 1, the Chair, Steven Willard noted there were two overarching questions the group sought to answer. The first question was "To what extent are the Merit Review process and criteria, as currently understood, implemented, and assessed resulting in awards for research and education that achieve NSF's mission?, while the second asks "Would changes to the Merit Review process and criteria, their implementation, or their assessment, are needed to achieve NSF's mission?" (Notably, this second question was read out to correct 'would' to 'what', likely indicating the ungrammatical phrasing was due to a typo on the slide.) These questions as presented to the group can be seen in Figure 11.

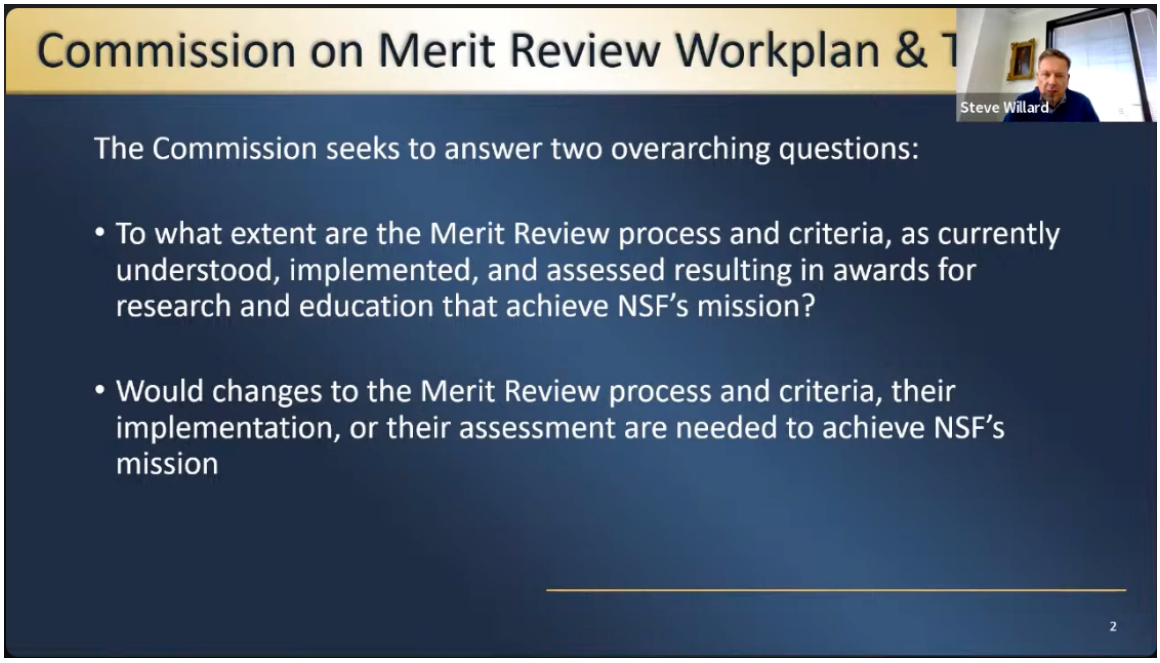

Figure 11. Screenshot of YouTube video with MRX Commission's guiding questions.

Both questions highlight three separate points of consideration - how the Merit Review process is understood, implemented, and assessed. These three aspects were formalized in the committee process as Policy, Implementation, and Accountability. Of the ten open MRX commission meetings analyzed, all followed a similar structure informed by these three aspects. The committee followed an approach that began with questions around policy, followed by implementation, and lastly accountability as described in the Workplan during MRX 1. This three-part structure can also be seen in the official meeting agendas for MRX 10 and 11, which focused on assessing and prioritizing potential policy recommendations and implementation and accountability recommendations respectively (Table 12). Unsurprisingly, the themes of policy, implementation, and accountability were at the forefront in MRX commission meetings as observed during analysis.

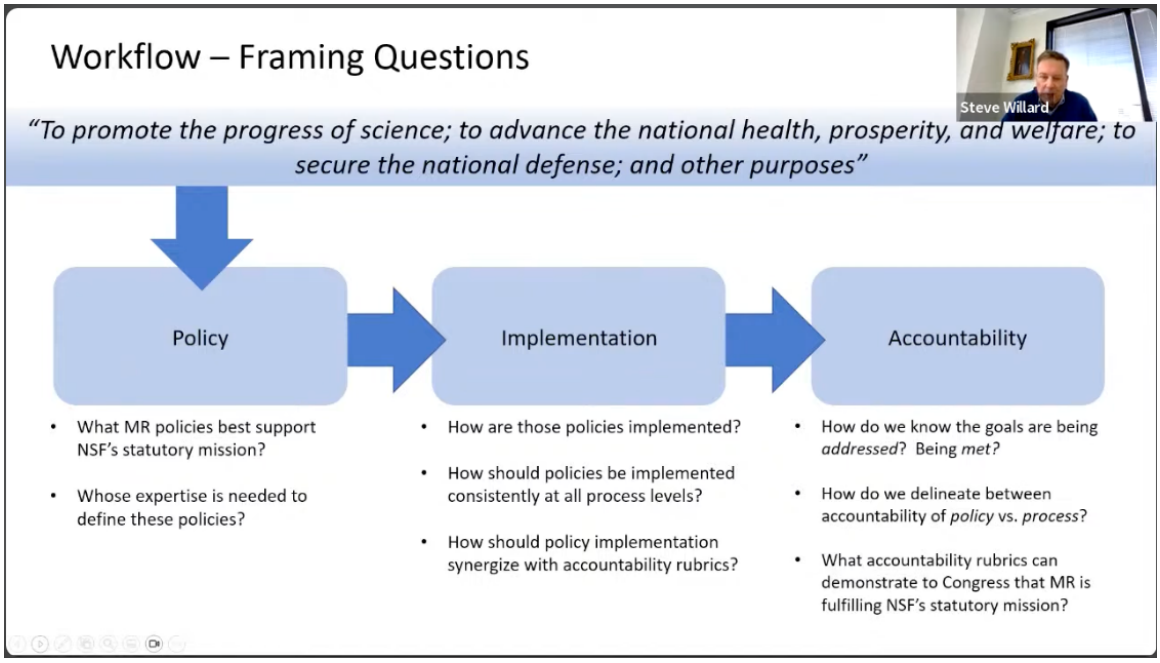

Figure 12. Screenshot of YouTube video showing framing questions for each of three focus areas.

Figure 12 shows the framing questions applied to each of the three aspects in the initial MRX committee workplan. While neatly laid out here, these three aspects of Policy, Implementation, and Accountability came to contain a much more complex range of concepts during the course of the MRX commission meetings. Emergent themes relating to policies around broader impacts, the interplay between policy and implementation within the NSF, and the role of government and institutional accountability in the NSF's decision-making processes came to shape the overall outcomes of the MRX Task Force as a whole.

4.3 Understanding Broader Impacts Policy

4.3.1 Weighting

With the introduction of Intellectual Merit and Broader Impacts as the revamped Merit Review criteria in 1997, the National Science Foundation quickly

found that that the question of criteria weighting between Intellectual Merit and Broader Impacts was a "central issue", with Intellectual Merit being strongly weighted as more important by public commenters (NSB-NSF Staff Merit Review Task Force 1997). Understanding how the two Merit Review criteria are weighted, and what policies govern those decisions was a significant point of concern for the Task Force:

> "So, for example, the weighting of criteria, I think that's a policy question that we've never really identified or discussed as a group. I mean, you have your intellectual merit, you have your broader impacts. How should those be weighted? So, I think maybe the question should be, is there a policy that dictates it?" (MRX 1)

To which another member, who at the time of the video served as a Program Director at the NSF, replied:

> "My understanding is that it's a holistic review that neither is weighted more than the other. They're considered simultaneously and in a complementary fashion. So, for example, if you were looking to uncover a policy that says we weigh intellectual merit this much and broader impacts this much, that's not going to be available." (MRX 1)

This statement, that the policy is not about weighting but rather 'considering' both criteria directly references the current NSF language in the Proposal Award Policies and Procedures Guide, which states that both criteria are due "Equal consideration" (National Science Foundation 2023), a fact that other Task Force members went on to reference in later meetings. Pushing back against the description of this approach as holistic review, as many members had been saying in discussions and as can be seen in the quote above, one member of the Task Force brought the discussion back to the language in the official Proposal and Award Policies and Procedures Guide (PAPPG), noting that the official policy in introducing the criteria was "Each criterion is necessary, neither in itself is sufficient, and full consideration

is not necessarily equal consideration" (MRX 7).

That full consideration is not necessarily the same as equal consideration was both intentional and a motivation for such language on the part of NSF, in response to broad public opinion from researchers and PIs, who found Intellectual Merit to be more important than Broader Impacts. This unequal weighting emphasizing Intellectual Merit over Broader Impacts was made clear not only during the reveal of the criteria in 1997, but also in the second 2011 Task Force on Merit Review. Discussing the history of criteria weighting and evaluation at the NSF, one Task Force member went on to describe how the 2011 Task Force continued to show the research community maintained those concerns around weighting:

> "Nearly 15 years later, a summary report of stakeholders' input for the 2011 NSB Task Force on Merit Review echoed similar yet nuanced concerns compared to those expressed by stakeholders in 1997. Most NSF staff, NSF Advisory Committee members, reviewers, and principal investigators believed that intellectual merit was weighted more than broader impact, and indeed found IM to be more important than BI. However, a larger proportion of NSF staff and advisory committee members, compared to reviewers and PIs, believed more equal weighting to be appropriate." (MRX 7)

This last, that a 'larger proportion of NSF staff' supported weighting Intellectual Merit and Broader Impacts more equally, references data from a survey conducted by the 2011 Task Force. A majority of survey respondents across the board found that Intellectual Merit had been weighted more than Broader Impacts in proposal review and selection. Moreover, those respondents agreed that that skew towards Intellectual Merit should be maintained in future decisions. In the 2011 survey, 87% of NSF officials expressed the view that Intellectual Merit was currently given greater weight in the proposal review and selection process. However, 69% of those officials indicated a desire to see a partial decrease in this emphasis, advocating for a more balanced approach while still maintaining some importance for

Intellectual Merit. In contrast, 68% of Principal Investigator (PI) respondents believed that Intellectual Merit was more heavily weighted in 2011, with 76% of PIs indicating a preference for Intellectual Merit to be weighted even more heavily than it was at that time (National Science Board 2011).

More recently than 2011, multiple MRX commission members reaffirmed their beliefs that Intellectual Merit continues to outweigh Broader Impacts, and that the NSF should maintain that, with statements from Task Force members such as:

> "I think intellectual merit should really be more than 50%. As important as broader impacts are, if we don't have high levels of intellectual merit, what's the point? I've come to some comfort level with them being equal, but I certainly don't want to see any erosion in any way, shape, or form of intellectual merit. " (MRX 10)

However, even with such a strong precedent for prioritizing Intellectual Merit over Broader Impacts, the discussion nevertheless returned repeatedly to the relative importance Task Force members did feel was owed to Broader Impacts. As one Task Force member concluded at the end of the session on Broader Impacts:

> "Part of it, I think, goes back to what Julia said at the beginning; it's about valuing both of the criteria equally. If you read what the principles and the merit review guidelines say, if you read them and take them to heart, it says you're supposed to do both of those things well, that they are the things that we judge proposals on, and we want them to do them well. Someone asked me in the last conversation whether it was okay to fund things that are okay in the science but really good in broader impacts, and I said absolutely not. What I want the NSF to fund is things that are excellent in broader impacts and excellent in intellectual merit." (MRX 3)

This quote from a Task Force member underscored the tensions faced among NSF staff between valuing Intellectual Merit and Broader Impacts equally in NSF proposal submissions, while also reaffirming the relatively greater import of Intellectual Merit in the review process. In the quote above, the MRX commission member advocated for placing significant effort into developing robust broader

impacts, asserting that both criteria had to be approached with equal seriousness and thoughtfulness by applicants.

However, this emphasis on *valuing* both criteria equally does not also imply that both criteria ought to be *weighted* equally in decision-making, as the quote made clear that "okay" science was insufficient, even if paired with excellent broader impacts. This remains an ongoing challenge in NSF's merit review process: ensuring that while broader impacts are taken seriously and thoughtfully considered, they do not overshadow the fundamental requirement for high-quality scientific inquiry.

With regard to the lingering question of how best to approach weighting Broader Impacts, the Task Force discussions reveal a persistent tension between Intellectual Merit and Broader Impacts in the NSF's Merit Review process. While official policy mandates 'equal consideration' of both criteria, the practice of prioritizing Intellectual Merit has been strongly supported by researchers and Principal Investigators, However, a significant proportion of NSF officials have also expressed a desire to shift towards a more balanced approach, advocating for a decrease in the emphasis on Intellectual Merit while still acknowledging its primary role (National Science Board 2011). This ongoing discrepancy between the perspectives of researchers and NSF staff underscores a broader institutional challenge for the NSF in aligning its Merit Review policies with its strategic goals.

4.2.3 Portfolio Balance

Another aspect that emerged during MRX commission discussions was understanding the NSF's approach to and policies around project portfolios and portfolio management. An aspect of the review process that is not often considered by applicants, portfolio balance highlights the NSF's necessary top-down, large-scale institutional perspective on Merit Review. This institutional perspective would come to be seen in codes such as 'Portfolio Balance' and 'Capacity Building', as seen in

Table 3. During the Task Force discussions, portfolios and portfolio balances were described as:

> "...basically a collection of projects which can be collected and coded on a variety of levels. Each director, division, and program have award portfolios, even individual Program Officers have award portfolios. Regardless, there are a variety of considerations NSF staff must keep in mind when building and managing a portfolio, which include intellectual diversity, risk diversity, institutional diversity, career diversity, demographic diversity, and geographic diversity. Program officers and division and directorate leadership must also develop both short and long-term visions for their portfolios, given the variation of proposal submissions over time. Internal priorities such as NSF's strategic plan, NSB's Vision 2030, and intellectual merit and broader impacts, along with external priorities such as legislation, executive orders, and Supreme Court rulings, all factor into proposal composition. However, NSF does not seem to have a formal definition of portfolio balance, which may be by design. After all, flexibility is important when responding to various goals." (MRX 8)

This flexible approach is understandable in light of the Foundation's broad mandate and national influence, a fact requiring it to consider the cumulative impact of its funding decisions on the advancement of knowledge and societal benefit, rather than solely the individual outcomes of specific research projects. However, the effects of this flexible, institutional perspective on portfolio balance become increasingly challenging in understanding the context of the NSF's policies with regard to balancing Intellectual Merit and Broader Impacts. As one Task Force member commented:

> "But in terms of the whole portfolio management, I don't know how that can be done at a proposal by proposal level. I mean, the broader impacts cover a really wide waterfront, and it is unrealistic, especially in the individual PI grants, to expect very much of that waterfront to be covered in any given proposal or project. I think it has to be up to the program manager or some group to look at the entire suite of proposals to ascertain that not every single one, in terms of broader impacts, is looking at, oh I don't know, K-3 STEM education, and nobody's paying any attention to the other aspects that go into broader impact. That would probably not be a good portfolio, any more than having all of the proposals in physics be focused on, say, condensed matter physics." (MRX 7)

This concern, that every single proposal might be considering the same or similar Broader Impacts space, is very real, as seen in Chapter 3 of this dissertation with concerns around the tendency to default to broadening participation through outreach to underrepresented minorities as fulfilling the Broader Impacts requirement. The breadth of potential Broader Impacts options in particular poses a unique challenge for the NSF with regard to portfolio balance. As the quote above notes, it is unrealistic to expect each individual proposal to comprehensively address all possible Broader Impacts within the scope of their project. Nor should individual PI's be expected to coordinate coverage of that breadth of possibilities. Instead, that responsibility falls to the Program Officers and Division Directors to oversee. This also emphasizes the unique role of NSF staff, in guiding and overseeing not just individual projects but also entire fields and disciplines. The absence of a formal definition or policy regarding portfolio balance, while allowing for this necessary flexibility, also allows that balance to be missed or overlooked, however, a fact that raises the issue of potential differences and inconsistencies in the balance and implementation the Merit Review criteria in actuality.

4.4 Implementing Merit Review On Paper and In Practice

4.4.1 Internal Culture at the NSF

The in vivo code 'Culture' was one raised repeatedly by members of the Task Force, particularly in a series of continuing conversations in MRX 2, and one that ranged across conversations of policy, implementation, and accountability alike. Commission members struggled with the overlap between policy and implementation, and particularly with how this blurred space was maintained by the internal culture of the NSF around the Merit Review process, both from reviewer panels and program awards.

Discussions of culture often revolved around the impact of that culture on how the Merit Review process plays out in theory and policy versus in actual implementation:

> "I wanted to connect though to something that [member] just said because it is true there's a real culture inside NSF. It applies in Academia as well because what we write down on paper about academic success for promotion and tenure, how we actually do it, they're not quite the same thing. And if you don't understand the unwritten lore, you will not be successful. And so, I do think as we talk about these issues, the words matter, but how we translate them into the operating culture really does matter, perhaps even more, because that will determine what happens in truth on the ground." (MRX 2)

This quote, which came at the beginning of a meeting focused on the current Merit Review criteria as set down in NSF guidelines, touched off a significant discussion of how the internal culture of the NSF in policy often differed from the so-called 'operating culture', as noted by one Task Force member in the quote above.

That gap in theory and in practice brought together questions of policy and implementation, and often came with significant ramifications. This can be seen in the conversation touched off by the above quote during a consideration of the five official Merit Review elements listed in the PAPPG. These five Merit Review elements, which are listed immediately after the definitions of Intellectual Merit and Broader Impacts, are a series of five questions that are to be considered during the review process, including "To what extent do the proposed activities suggest and explore creative, original, or potentially transformative concepts?" And "How well qualified is the individual, team, or organization to conduct the proposed work?" (National Science Foundation 2023, III.A.2).

In reviewing external resources on grant preparation and submission, these five elements are not commonly included in most external resources, particularly not those dedicated to Broader Impacts support, such as those offered by universities (University of Wisconsin-Madison 2023, Virginia Tech 2024), most of which focus on

the definition of Broader Impacts immediately preceding the five elements in the PAPPG (2023, III.A.2) and the list of potential outcomes provided by the NSF (II.D.2.d.i). The questions contained within the five Merit Review elements do appear in some external resources (Center for Advancing Research in Society 2023, MIT School of Science 2024), but are often presented as piecemeal questions rather than specifically as the five Merit Review elements.

The relative absence of these five elements in external discussions on grant preparation and submission, coupled with their general lack of awareness and integration within broader discourse, was noted by Task Force members during the presentation of the five elements during MRX 2. As one Task Force member asked,

> "What I'm getting at is that, right now, we're talking about whether we like the merit review elements, and it's making me ask, at the end of the day, whether we liked them or not, the question is, to what extent do they drive the very structure of the review process?" (MRX 2)

To which a guest speaker and National Science Board Chair responded:

> "Honest answer: not very much. I mean, I think it's a practice, and again, this is a couple of years ago [...] but people show up with some written comments, there's a debate, but I don't ever remember us going through the criteria and saying, you know "what do we think about these?". It's exactly what you said, here's the intellectual merit, are we satisfied with the broader impacts? Okay, now let's argue about the proposals in the pool." (MRX 2)

While one other Task Force member commented to say they had used the five Merit Review elements specifically in review processes they had participated in and also conducted themselves as a Program Officer, the majority of Task Force members agreed that they had seen and experienced minimal implementation of the elements themselves, despite their role as stated during the Task Force proceedings as the primary framework for the review process, meant to "provide guidance towards ultimately achieving the NSF's mission" and "help clarify consideration of the

criteria" (MRX 2). This also echoes the relative lack of emphasis of the Merit Review elements in external resources and the broader conversation around the Merit Review criteria.

This discrepancy between formal policy guidelines and practical application during the review process was summed up by the same member who raised the question as follows:

> I would say, though, this goes back to our earlier conversation about culture. There's the de facto and there's the de jure, and they're not the same thing. So, if we want a change in the review process, we actually have to reshape the culture. It's probably more important than the words. The words are the supporting evidence for the culture you want to drive, but it's the culture that determines what happens. (MRX 2)

Here, the discussion of internal NSF culture, again coded as 'Culture', was explicitly highlighted as not only varying in implementation, but also as something that could - and potentially should - be changed in service to improving the Merit Review process.

4.4.2 Consistency and Flexibility in Implementation

The topics of consistency & flexibility were a significant aspect of the discussion on implementation, which often ranged in scope from the internal implementation of the criteria within the NSF, to the implementation of the criteria by institutions and researchers on a more individual level. Within the NSF, consistency in the implementation of Broader Impacts across the various divisions and directorates was a significant point of discussion during MRX commission sessions.

> "One of them that I hadn't really thought about before, that [the Task Force Chair] brought up, was the difference across directorates and, I imagine, probably even across programs, in terms of the way in which broader impact is interpreted and potentially even the seriousness with which it is taken. I think that is a very high priority for the foundation to deal with. You know, it's known as walking the talk, right? So, that is at the foundation level." (MRX 3)

One aspect of the discussion around the implementation of the Merit Review criteria generally, and Broader Impacts specifically, was the role and value of flexibility in both proposals and research largely. The National Science Foundation has maintained the value of flexibility in the proposal review and selection process since its founding, and that was echoed during MRX commission discussions:

> "But I want to go back to one comment I heard someone say around the differences the way BI is implemented across the directorates. This is an important area that I think we want to talk about as we talk about implementation. Part of what is the value of the criteria is that it sets up for all of NSF to see the criteria and add flexibility for various disciplines, various projects, various investigators who have very different experiences, different resources to still be able to do broader impacts, even if they're not in an R1 institution, or that they don't have an Institutional Broader Impact Center, or that they're an early-career faculty member. The policy or the way in which we set up the criteria is really about allowing people to access the merit review process through intellectual merit and broader impacts. The way in which disciplines build up their own perspectives and experiences around broader impacts, in my opinion, should remain a disciplinary opportunity. We shouldn't make everyone try to do broader impacts the exact same way across all disciplines. We don't do that for science; we shouldn't do it for broader impacts." (MRX 3)

In the context of the current Merit Review criteria, the NSF has specifically emphasized the value of maintaining flexibility in Broader Impacts, as seen in the evolution of guidance provided around the criteria from 1997 through to the time of this third Task Force in 2023, and the organizations hesitance to provide increased guidelines on Broader Impacts out of concern those potential topics might be perceived as mandates or a 'checklist' (National Science Board 2011). The above quote highlights how the NSF understands flexibility in the context of the Broader

Impacts criterion, specifically by increasing accessibility and as a means of allowing various disciplines a greater range of opportunities and options.

However, what the NSF sees as valuable flexibility, researchers and PIs more often interpret as vagueness resulting in inconsistency. This can be seen in discussion of the results of a listening session on the Merit Review criteria hosted by the Center for Advancing Research in Society (ARIS) on behalf of the MRX commission:

> "Another prominent theme was guidance and training. Participants suggested that current guidance for BI project development is vague. They stated a need for better guidance to proposers and reviewers and more BI expertise on review panels, as we mentioned a moment ago. [...] Furthermore, several participants noted inconsistencies across the agency, that is, NSF itself, resulting in the perception that certain NSF programs or directorates prioritize BI more than others. Participants were interested in understanding how BI funding and outcomes differ by directorate. (MRX 3)

As seen in the above quote, researchers and PIs often perceive this flexibility as vagueness, or a lack of clear, concrete guidelines, leaving them without a clear path to meeting the requirements for funding. However, by contrast, the NSF views flexibility as essential for providing accessible and attainable broader impacts opportunities, accommodating diverse research contexts. Given this divergence, there is an apparent disconnect between the National Science Foundation's (NSF) interpretation of flexibility in the Merit Review process and how it is perceived by researchers and Principal Investigators (PIs).

4.5 Accountability

4.5.1 Governmental Accountability

Accountability was yet another theme that emerged in the discussions. MRX commission members were keenly aware of the specter of government accountability in the decision-making process, and of the role that government oversight played in shaping internal NSF culture, particularly around high-risk or transformative

research. Coded as “Government Accountability’, this topic typically was raised in conjunction with proposed changes to or modifications of existing NSF guidelines, or either of the two Merit Review criteria, typically with regard to concerns or hesitations to acting too quickly or without enough consideration of legislative constraints:

> “…the legislation is telling us this is what you'll do. And somewhere in there too is, and you'll report on it. So, we need to be a bit careful that we're reconciling for our discussions and then back into what we're required to do by the legislation.” (MRX 2)

However, hearkening back to the “Conlan Wars” and the increased Congressional and public scrutiny the NSF faced beginning in the 1970s, the historical background underpinning these concerns with government accountability was also recognized:

> “We have to find ways to encourage and empower, and this speaks to the culture. I always remind ourselves that there are some real reasons that there has been a rather conservative approach historically because when something goes wrong, it's big and it's all over the newspaper, and you can rest assured we're going to the hill. That's why Merit Review is one of the issues that Congress constantly takes a look at and oversees.” (MRX 8)

This reflects concerns raised in the article “Science Funders, High-Risk Research, And Apple Pie” by the Good Science Project (2023), which was discussed by the MRX Commission. MRX commission members largely agreed with the general thesis of the article, which speaks to the challenge of changing the culture at the NSF with regard to ‘empowering’ Program Officers to fund high-risk science absent the traditional peer review approach. The article elaborates on so-called ‘apple pie positions’, where seemingly simple assertions (i.e. “we should wait until we gather more data”) are challenging to question and difficult to publicly criticize, even when they have perverse effects on the process at hand, particularly hesitation to change or act.

In this case, with the process being scientific funding and grant awards, the ability of the NSF as an organization to act quickly and decisively becomes bogged down by apple pie positions that hamper Program Officers and Division Directors from making high-risk funding decisions. Even with the challenge of 'apple pie positions' recognized, and with the majority of commission members in support of funding more high-risk, transformative research, adding government accountability and oversight to that mix primarily led to hesitation.

4.5.2 Institutional accountability and support/buy-in

Accountability, from the National Science Foundation to Congress and other Governmental bodies shaped much of the discussion on the topic, particularly with regard to evaluation of Broader Impacts, an issue raised to prominence in the Chips and Science Act. However, MRX commission members also saw accountability flowing the other direction, from researchers, PIs, research centers, and institutions of higher education, to the NSF itself. In discussion of the results of the listening session hosted by ARIS on behalf of the Task Force, it was noted that:

> "With regard to institutional support for BI, this was another emerging theme, with participants expressing varying degrees of support for BI from their institutions. They suggested that institutions of higher education do an inadequate job, if at all, at tracking and documenting the impact of BI, whether at the PI, the departmental, or the institutional level. This lack of tracking was perceived to be connected to a downplaying of societal impacts in promotion and tenure packages. There was also a general opinion that NSF does not sufficiently hold institutions accountable for their institutional capacity building around BI." (MRX 3)

Institutional support was directly tied to concerns around the weighting of Broader Impacts and Intellectual merit. This concern around institutional 'buy-in' on Broader Impacts as whole was a fundamental issue on which researchers and PIs sought guidance and support from the NSF in holding their colleagues and own

institutions accountable for funding, supporting, and rewarding Broader Impacts to the same level as Intellectual Merit.

MRX Commission members concurred with these concerns, and repeatedly considered issues related to this challenge, discussions of which were coded under "Institutional Support". These debates ultimately described issues of institutional support that lay both with the administration and culture of research centers and institutions of higher education broadly, but also with the long term effects of minimal training in and low-buy in on Broader Impacts among individual researchers as well. Speaking as a former Program Officer and Division Director, one Task Force member noted:

> "I think what a lot of what we're dealing with is the fact that you have people who come through the entire educational system, they find themselves in a tenure-track position, and suddenly they are writing their own proposals, and they have to do this thing called broader impacts that most of them have had minimal exposure to in their entire educational career. And so, between the fact that you have a population of people writing these proposals who don't really know much about this aspect and they're very comfortable with the science and proposing that, so there's a huge disconnect there. And the fact that many of the institutions at least give the impression, according to what Wanda said, and it's quite believable to me, that many of the institutions are sort of ambivalent about it, or at least will definitely put the weighting over on the science and the research side, as opposed to the broader impact side. So, you don't have a culture that values this or even has a clue of where to start unless you dive in." (MRX 3)

While no solutions were offered in discussions, nor have significant recommendations yet been made concerning improving institutional accountability for Broader Impacts, the topic was ultimately recognized as a key issue among researchers and PIs that requires addressing. Options will likely be discussed more formally when final recommendations are announced.

5. Generalizability and Limitations

While the MRX commission open meeting recordings transcribed and analyzed in this work provided a valuable window into the inner decision-making processes shaping a major scientific funder, those meetings were also only a portion of the work being done by the Task Force. As closed meetings or meeting notes were not publicly available, part of the picture contained in those closed-door discussions cannot be accounted for in this analysis. As such, the results should be taken as a limited perspective on the internal workings of the National Science Foundation's interpretation and implementation of the Merit Review criteria. However, given the paucity of information from this perspective, the results are nevertheless valuable, even if limited.

Furthermore, this Task Force was only one of several efforts the NSF has held throughout the decades, as the Merit Review criteria continue to evolve, and the NSF itself is only one of many scientific funding organizations in the US and North America more broadly. As such, these findings are necessarily a snapshot of the organization's approach at this moment in time and cannot reasonably be generalized outside of that context.

6. Discussion and Conclusions

A novel LLM-assisted transcription process was used to clean transcripts of the NSB-NSF Task Force On Merit Review for analysis. The process was overall successful, and greatly simplified the arduous and often tedious process of transcribing and cleaning hours of meeting notes. That this material, ostensibly covering public meeting notes of a matter of federal importance, was both so difficult to locate and challenging to analyze in its original audio-visual format offers a valuable opportunity for using similar LLM-assisted transcription methods to increase access to and visibility of similar public content.

It is important to note that given the private status of many LLM companies and the opaque nature of data usage by ChatGPT and similar systems, the author finds that the primary use of this method should remain focused on publicly available data. The application of LLM-assisted transcription in this study also demonstrated the potential for more efficient and accurate data collection and cleaning of public committee data, though challenges may remain in transcribing less structured speech and more research is needed to understand the strengths and limitations of this method with more variable types of data inputs.

Considering the data itself overall, the discussions held in the ongoing NSB-NSF Task Force on Merit Review reflected the persistent challenges the Foundation faces in balancing Intellectual Merit and Broader Impacts within the NSF's current Merit Review process. Those concerns continued to echo issues raised much earlier in the evolution of the Merit Review criteria and proposal selection process, even as their current forms may look new. Weighting remained a central policy issue for both Task Force members and researchers and PIs, as it has been in previous discussions held by the NSF and NSB. Even with formal guidance calling for equal consideration of both criteria, both past and present practice has favored Intellectual Merit, as evidenced by public comments and survey data from the 1997 and 2011 Task Forces (NSB-NSF Staff Merit Review Task Force 1997; National Science Board 2011).

Yet, despite the strong precedent for prioritizing Intellectual Merit, Task Force debates frequently revisited the importance of Broader Impacts. While researchers generally favored Intellectual Merit, many NSF officials and Task Force members were actively working to identify a more balanced approach that the broader scientific community might accept, without undermining Intellectual Merit. All of these factors reflected the NSF's ongoing institutional challenge in aligning Merit Review policies with the organizations' strategic goals. In this sense, the 2023 Task Force's deliberations reasonably reflect the complexity of implementing a review

process that not only adheres to official policy but also responds to the broader scientific community's priorities.

In considering the guiding theme of Implementation, debates and questions often blurred the line between policy and implementation; overall matters of implementation were largely tied to questions around policy in theory and in practice, as seen with the official policy versus minimal real implementation of the five Merit Review elements. Task Force discussions also noted the tensions evident between consistency and flexibility in the implementation of Broader Impacts. While the NSF has historically valued flexibility and with Broader Impacts, seen flexibility as a means of widening options and increasing institutional and individual means of addressing the Broader Impacts criterion.

However, in contrast, researchers have often interpreted this flexibility as vagueness. This miscommunication has historically led to frustration and concern around potential inconsistent application of Broader Impacts across divisions. Moving forward, the NSF must strike a careful balance between maintaining flexibility, which it values for promoting inclusivity, and providing the clarity that researchers seek to meet funding requirements. Particularly as the NSF looks to increase evaluation of Broader Impact as requested by the CHIPS and Science Act, addressing this tension will be essential for ensuring that the Merit Review process both aligns with the NSF's goals and meets the needs of the research community

Accountability remained the third guiding theme within the analysis, and one that played a significant role in MRX commission discussions. Comments around accountability considered both on behalf of individual researchers and institutions towards the NSF, as well as the NSF's accountability to Congress and the government more broadly. Concerns were raised about the lack of institutional support for Broader Impacts, with researchers seeking more guidance and support from the NSF

in holding research institutions accountable, to ensure Broader Impacts receive equal recognition and resources as Intellectual Merit.

The NSF also remained concerned with its own accountability to Congress, which for the Task Force primarily takes the shape of two preliminary reports, scheduled for May and July 2024, and a final report on the Merit Review process expected in December 2024. It is important to note that the final recommendations by the Task Force are not yet available, and are not expected until late 2024. As of the time of writing, the NSF had issued a public Request for Information (RFI) around the Merit Review criteria. Data collection in that effort was ongoing in Fall 2024, and expected to be seen in the final report still scheduled to be delivered in late 2024.

Additionally, the interim report, containing a "preliminary set of recommendations for public comment" (National Science Board 2024), which was supposed to be shared in May 2024 and was publicly referenced in MRX 10, 11, and the Plenary talk, could not be identified in the public record as of the time of writing. An overview of the preliminary recommendations, however, was provided in the plenary talk delivered by the Task Force Chair. Those preliminary policy recommendations, which included that "NSB should reaffirm that NSF's Merit Review is the gold standard for awarding funding", "the two current criteria, Intellectual Merit and Broader Impacts, fully meet NSF's mission", and that "that NSF continues to promote the progress of science in accordance with the NSF Act and other applicable statutes" (MRX Plenary). Taken in conjunction with the NSF's noted concerns around government accountability do not represent significant deviations from previous NSF policies on Merit Review.

Therefore, while the exact final recommendations are yet to be known, it is also reasonable to conclude that much like many policy-focused projects, changes and progress can be expected to be incremental. How those final recommendations and the data collected in the RFI might change the landscape of Merit Review

remains to be seen. Yet, considering the ways in which the history of the review process and decision-making within the National Science Foundation and previous Merit Review Task Forces has incrementally built on both the internal culture of the Foundation and the evolving interests, needs, and values of the research community, a snapshot of how the NSF understands the Merit Review criteria at this moment in history nevertheless offers valuable insight into where the future of scientific funding in the United States might be headed.

7. References

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, *15*(2).

Center for Advancing Research in Society. (2023). ARIS Broader Impacts Toolkit. Advancing Research in Society. https://aris.marine.rutgers.edu/

Cotos, E. (2019). Articulating societal benefits in grant proposals: Move analysis of Broader Impacts. *English for Specific Purposes*, *54*, 15-34.

Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research*, 50(2), 708-739.

Frodeman, Robert, J. Britt Holbrook, Patricia S. Bourexis, Susan B. Cook, Laura Diederick, and Richard A. Tankersley. "Broader Impacts 2.0: seeing—and Seizing—the opportunity." *BioScience* 63, no. 3 (2013): 153-154.

Good Science Project. (2023, July 18). Science Funders, High-Risk Research, and Apple Pie. https://goodscienceproject.org/articles/science-funders-high-risk-research-and-apple-pie/

Google Support. (2024). *Translation & transcription glossary—YouTube Help*. YouTube Help. https://support.google.com/youtube/answer/7296221?hl=en&sjid=3459571861152418541-NC

Google Support. (2024b). *Use automatic captioning—YouTube Help*. YouTube Help. https://support.google.com/youtube/answer/6373554?hl=en

Gould, R. K., Coleman, K. J., Krymkowski, D. H., Zafira, I., Gibbs-Plessl, T., & Doty, A. (2019). Broader impacts in conservation research. *Conservation Science and Practice*, *1*(11), e108.

Hallinen, J. R. (2014). The many quiet tensions: Perceptions of the broader impacts criterion held by NSF career award holders at very high research institutions of higher education.

Heath, K. D., Bagley, E., Berkey, A. J., Birlenbach, D. M., Carr-Markell, M. K., Crawford, J. W., ... & Wesseln, C. J. (2014). Amplify the signal: Graduate training in broader impacts of scientific research. *BioScience*, *64*(6), 517-523.

Kamenetzky, J. R. (2013). Opportunities for impact: Statistical analysis of the National Science Foundation's broader impacts criterion. *Science and Public Policy*, *40*(1), 72-84.

Mazuzan, G. T. (1992). " Good Science Gets Funded..." The Historical Evolution of Grant Making at the National Science Foundation. Knowledge, 14(1), 63-90.

MIT School of Science. (2024, January). NSF Broader Impacts statement guidance. MIT School of Science. https://science.mit.edu/resource/nsf-broader-impacts-statement-guidance/

Nagy, D. (2016). Determinants of Broader Impacts Activities: A Survey of NSF-Funded Investigators. *Journal of Research Administration*, *47*(1), 68-93.Nagy, D. (2016). Determinants of Broader Impacts Activities: A Survey of NSF-Funded Investigators. *Journal of Research Administration*, *47*(1), 68-93.

National Alliance for Broader Impacts (NABI). (2018) The current state of broader impacts: advancing science and benefiting society. Columbia (MO): National Alliance for Broader Impacts. https://www.broaderimpacts.net

National Science Board. (2011). National Science Foundation's Merit Review Criteria: Review and Recommendations (NSB/MR-11-22). National Science Board. https://www.nsf.gov/pubs/issuances/in121.txt

National Science Board. (2023). *NSB-NSF Commission on Merit Review (MRX)*. National Science Board. http://www.nsf.gov/nsb

National Science Board. (2024). *Merit Review Reexamination Commission Resources*. National Science Board. http://www.nsf.gov/nsb/NSBActivities/merit-review-commission.jsp

National Science Foundation (2023). *Proposal and Award Policies and Procedures Guide* (NSF 24-1, OMB Control Number 3145-0058). Washington, DC.

NSB-NSF Staff Merit Review Task Force. (1997, March). Merit Review Final Recommendations. https://www.nsf.gov/pubs/1997/nsbmr975/nsbmr975.htm

OpenAI. (2024). ChatGPT (4.0) [Large language model]. https://chat.openai.com/

Reilly, C. A. (2021). Searching for the “magic elixir”: Uncertainty in developing specific aims and broader impacts for NIH and NSF grants. *The Routledge Handbook of Scientific Communication*, 189-199.

Roberts, M. R. (2009). Realizing societal benefit from academic research: Analysis of the National Science Foundation’s broader impacts criterion. *Social Epistemology*, *23*(3-4), 199-219.

Rothenberg, M. (2010). Making judgments about grant proposals: A brief history of the merit review criteria at the National Science Foundation. Technology & Innovation, 12(3), 189-195.

Siontis, K. C., Attia, Z. I., Asirvatham, S. J., & Friedman, P. A. (2024). ChatGPT hallucinating: can it get any more humanlike?.

Skrip, M. M. (2015). Crafting and evaluating Broader Impact activities: a theory-based guide for scientists. *Frontiers in Ecology and the Environment*, 13(5), 273-279.

University of Wisconsin-Madison. (2024). *Broader Impacts Homepage*. University of Wisconsin-Madison. Retrieved August 19, 2024, from https://broaderimpacts.wisc.edu/

Virginia Tech. (2024, July). *Broader Impacts Partnerships*. Center for Educational Networks and Impacts. https://ceni.icat.vt.edu/content/ceni_icat_vt_edu/en/partnership-process.html

Wiley, S. L. (2014). Doing broader impacts? the National Science Foundation (NSF) broader impacts criterion and communication-based activities.

CHAPTER 5

CONCLUSION

This dissertation set out to examine the complex processes by which the National Science Foundation's (NSF) Merit Review criteria are understood and implemented by various stakeholders, including individual researchers, Program Officers, and institutional officials within the NSF. Through a combination of survey data, semi-structured interviews, and public meeting notes from the NSF-NSB Task Force on Merit Review, this research sought to provide a detailed exploration of the differing perspectives and tensions that shape the proposal submission and selection process.

The findings of the three papers in this dissertation, encompassing Chapters 2-4, highlight the important role of the NSF in influencing funding practices nationwide, but also how individual researchers' needs and concerns are shaping the organization from the bottom up in turn. While various aspects of both the individual and institutional perspective were considered largely independently in previous chapters, it is also worthwhile to consider how those perspectives converge and diverge. Integrating all three papers together, the ways in which the individual researcher and institutional perspectives of Merit Review and the Broader Impacts criterion align - or do not - may help guide future research on this topic.

1. Synthesis of Key Findings

In Chapter 2, the focus was on the enduring concerns regarding the weighting of Intellectual Merit versus Broader Impacts in the National Science Foundation's (NSF) merit review process, which generally found Intellectual Merit to be more strongly weighted by both researchers and Program Officers. Questions around this

issue, initially raised during the introduction of the Broader Impacts and Intellectual Merit criteria in 1997, remain unresolved in 2024.

Moving on, Chapter 3 considered the range of Broader Impacts activities executed by surveyed researchers, and also began to bridge the gap between the individual and institutional by shifting perspectives from the individual researcher, to Program Officer's more institutionally-informed perspectives on larger trends in Broader Impacts activities, and then further to the emergence of Broader Impacts Centers and other institutionalized Broader Impacts efforts.

Lastly, Chapter 4 considered the National Science Foundation's institutional perspective through open meeting notes from the 2023-2024 NSB-NSF Task Force on Merit Review, which revealed an enduring emphasis on Intellectual Merit. The Task Force discussions underscored the complexity of implementing a Merit Review process that was both consistent and flexible, with accountability emerging as a central concern as the NSF looks to the future while navigating its obligations to Congress and the scientific community.

Following the Individual-Institutional axis along which these papers were initially structured (Figure 1), a second axis developed, that describes a series of enduring and emerging concerns. Here, enduring concerns can be considered those such as weighting, which has an established historical precedent dating back to the introduction of the revised Merit Review criteria in 1997 (NSB-NSF Staff Merit Review Task Force, 1997) and yet remains a point of debate in 2024. Other points, such as the development of Broader Impacts centers and questions around the value and consequences of siloing Broader Impacts efforts away from researchers into the hands of dedicated professionals, were not foreseen in previous Task Forces and thus represent emerging concerns that must be considered in the current context of research funding in the United States. As such, Figure 1 can be updated:

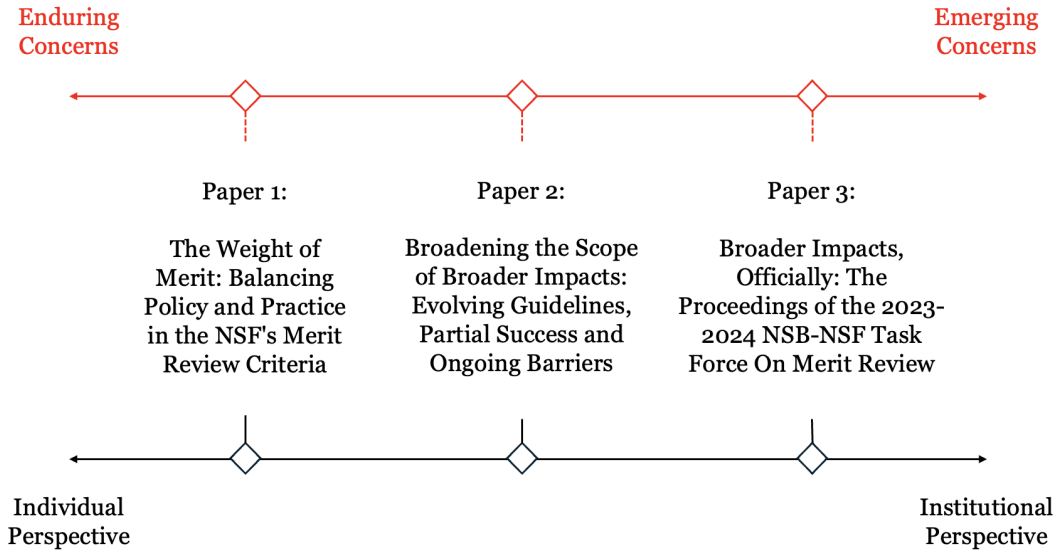

Figure 13. Update to Figure 1, showing the position of each paper in this dissertation along an axis of individual-institutional perspective, and along a scale of enduring-to-emerging concerns.

Following this, findings from all three papers can be integrated to consider how those emerging and enduring concerns have shaped researchers, Program Officers, and institutional officials' understandings of the Merit Review criteria, and of Broader Impacts specifically.

2. Weighting: An Enduring Concern for All

Despite the NSF's mandate for "equal consideration" of both criteria (National Science Foundation, 2023), the interview data revealed that researchers and Program Officers perceived Intellectual Merit as outweighing Broader Impacts, both in proposal reviews and in the greater context of career advancement. While the majority of researchers regarded Broader Impacts as important, many still internally prioritized Intellectual Merit more heavily, a matter which elicited some tension among researchers. This tendency stemmed both from their personal values concerning research and from a perception that the proposal review process and the

National Science Foundation as a whole placed greater weight on Intellectual Merit. Program officers echoed these concerns, noting that the NSF's policies and expectations enforce a schema in which Intellectual Merit is prioritized over Broader Impacts, and in some cases leaving researchers feeling 'compelled' to align with perceived NSF weighting of the criteria in order to secure grants.

Moreover, results from the open meetings of the NSB-NSF Task Force on Merit Review indicated that weighting also remained a significant and enduring concern for the Foundation. NSF officials and Task Force Members' statements and opinions weighted Intellectual Merit over Broader Impacts. In doing so, the institutional perspective on weighting aligned with researchers and former Program Officers' perceptions that Intellectual Merit was - and should be - more heavily weighted versus Broader Impacts.

Similarly to the survey results among researchers though, Task Force members also acknowledged Broader Impacts as important, though not as important as Intellectual Merit. To this end, Task Force members advocated for policies supporting a more balanced, albeit still unequal approach, to weighting the two criteria. In this sense, while weighting remains an enduring concern amongst all parties, those parties' understandings of the Merit Review criteria and perceptions of how those criteria are interpreted and weighted by all actors are in general alignment, leaving only the question of what should be done about addressing those concerns.

3. Accountability for Tenure and Promotion

As the implications of this imbalanced weighting extend beyond the proposal review process to the promotion and tenure process, it is clear that those concerns must be addressed. Both researchers and Program Officers reported that Broader Impacts efforts were often undervalued in comparison to Intellectual Merit when it came to promotion and tenure discussions. This was an 'unspoken' reality that Task

Force members were also aware of. This perception suggests that the current evaluation frameworks within institutions do not fully recognize or reward the societal contributions that the Broader Impacts criterion is intended to promote, leading to calls amongst researchers for increased accountability on behalf of their home institutions to the National Science Foundation in regards to incentivizing and rewarding Broader Impacts work.

This question of institutional support among universities and research centers was another place where individual researchers' and institutional officials' concerns aligned, with Task Force members acknowledging the issue and the need for improvements as seen in Chapter 4. Citing discrepancies in policy and practice within the internal culture of the National Science Foundation as a contributing factor cascading down to affect tenure and promotion processes, the MRX commission also faced the emerging issue around how to address those variations internally without losing the flexibility intentionally built into the Merit Review process.

4. Flexibility and Inconsistency as Dueling Concerns

The question of flexibility as understood by the institutional officials at the National Science Foundation was tied to not only internal culture at the Foundation, but also to its policies and expectations regarding the scope of Broader Impacts. Accomplishing a breadth of Broader Impacts through the criterion represents an enduring concern for the Foundation that was first raised with the introduction of the criteria (NSB-NSF Staff Merit Review Task Force, 1997).

At that time, one aim of the first Task Force was craft a Broader Impacts criterion that would allow for - and even encourage - innovation and flexibility in potential Broader Impacts approaches in order to maximize societal benefit (Rothenberg, 2010). Survey responses shown in Paper 2 (Chapter 3) indicate that the

National Science Foundation's flexible approach to the Broader Impacts criterion has successfully led to a diverse range of Broader Impacts activities among researchers.

Yet, researchers often expressed frustration with the Merit Review criteria, interpreting the intentional flexibility offered by the NSF guidelines as vagueness and seeking clarification or direction on which Broader Impacts activities are likeliest to secure funding or easiest to do as part of a research project, as seen in Chapter 3. Trends among these activities, particularly a strong co-occurrence between Broadening Participation and Training and Education shown in Chapter 3, show that these types were potentially favored by researchers for their ability to integrate comfortably with the familiar paradigm of research while directly contributing to a project's goals through student research work.

However, this tendency also led to concerns from Program Officers, who observed that many researchers approached Broadening Participation and Training and Education type Broader Impacts as default solutions rather than integral parts of projects' core research objectives. The distinction between the self-termed 'birds-eye-view' perspective of these former Program Officers and the narrower focus of individual researchers shows the shift from individual to institutional perspectives can have material effects on each party's understandings of the Merit Review criteria, and of Broader Impacts specifically. This 'clumping' of Broader Impacts types also show how a strictly-focused individual perspective on Broader Impacts can lead to an uneven impact in practice, an unintended consequence which institutional or larger-scale perspectives can help mitigate.

5. Portfolio Balance and Impact Identities: Emerging Approaches

Chapter 4 showcased some insights into the NSF's institutional perspective on the Merit Review criteria, particularly regarding the management of project portfolios and the concept of 'Portfolio Balance', introducing an institutional

perspective on the necessity of balancing approaches to Broader Impacts at scale. Portfolio balance was unique to the institutional perspective as, unlike individual researchers who often focus on the specific outcomes of their projects, the NSF must consider the broader, cumulative impact of its funding decisions. This institutional perspective necessitates a flexible approach to portfolio balance, ensuring that the overall body of funded projects advances both knowledge and societal benefit across the entire range of projects within a division, rather than merely on a project-by-project basis. The NSF's approach in particular requires Program Officers and Division Directors to manage portfolio balance, ensuring diverse and meaningful Broader Impacts across projects.

Similar to how the concept of portfolio balance requires NSF officials to take a long-term, holistic approach to managing projects within their divisions, some institutions are now encouraging researchers to cultivate an "Impact Identity." This concept, introduced by Risien and Storksdieck (2018), suggests that a researcher's impact, including their Broader Impacts activities, should be viewed as a continuously evolving process throughout their entire career arc, rather than being addressed on a project-by-project basis. This approach promotes a sustained commitment to societal impact, integrating Broader Impacts into the researcher's professional identity over time, rather than treating such efforts as isolated obligations tied to individual grants. In this sense, longitudinal approaches to research impact and Broader Impacts work more closely echo the work researchers do develop a traditional research direction and portfolio over the course of their careers.

A series of third-party efforts, including a workshop on "Unveiling Your Impact Identity" hosted by the National Alliance for Broader Impacts (2018), and another hosted by Michigan State University (Smith, 2022) on "Building Your Research Impact Identity and Vision" join resources such as Cornell's facilitated

exercises for researchers on Impact Identities (Fleming, n.d.) to encourage and support researchers in developing a long-term professional identity around research impact. Furthermore, such efforts may help scholars and researchers avoid defaulting to generic Broader Impacts activities such as bringing underrepresented minorities into a research project without long-term consideration of that impact work.

6. Individual or Institutionalized Impact?

While efforts to develop Impact Identities ask individual researchers to be responsible for their own understandings of impact and the development of impactful research work, other efforts question if that is the most effective way to approach research impact at all. In addition to general types of Broader Impacts, Chapter 3 also raised an emerging concern around the development of Broader Impacts centers and other forms of institutionalized Broader Impacts, such as networks and repositories. These efforts have become increasingly visible as, in response to the Broader Impacts criterion, the US research community has begun to develop a population of focused resources and dedicated professionals aimed at maximizing the efficiency of Broader Impacts projects and their associated funding lines.

The establishment of institutionalized Broader Impacts efforts through dedicated centers and institutes presents a compelling argument, as they have the potential to enhance the efficiency of Broader Impacts activities, maximize researchers' efforts, and provide crucial support to those who are hesitant or inexperienced in this area. However, the results of this dissertation indicate that opinions on such efforts from researchers and former Program Officers were mixed, with some seeking the greater partnership and support, while others worried that 'BI Centers' might ultimately distance researchers from impact entirely in service of more efficient impact outcomes. Finding a balance between supporting established

and dedicated Broader Impacts resources at the university level while still encouraging researchers to value and develop individual Impact Identities is a long-term challenge that will require balancing - and potentially redefining - not only Broader Impacts but Intellectual Merit in a way that is both understood and supported by not only the National Science Foundation, but individual researchers and their home institutions and universities together.

7. Final Considerations

Considering all three of these papers together, this research has begun to explore the complexities surrounding the National Science Foundation's Merit Review criteria, and particularly the Broader Impacts criterion, from both individual and institutional perspectives. The surveys, interviews, and public meeting notes presented reveal an enduring imbalance between Intellectual Merit and Broader Impacts, and the tensions both individual researchers and institutional officials face as they personal and public research needs, values, and practicalities. This imbalance, rooted in both individual and institutional perspectives, underscores the ongoing challenges faced by researchers and Program Officers who feel compelled both internally and externally to prioritize Intellectual Merit due to its perceived greater influence on funding and professional progression, as well as a desire to prioritize 'good science'.

The findings of this dissertation also revealed a critical need for enhanced guidance and support from both the NSF and academic institutions to help researchers more effectively integrate Broader Impacts into their work. While there is broad approval of the Broader Impacts criterion, the difficulty lies in its practical implementation, which is often seen as additional to or on top of the primary research work with Intellectual Merit, particularly due to a lack of incentive and reward from universities and research centers for strong Broader Impacts work. This

research advocates for a more equitable consideration and reward for efforts related to both criteria, suggesting that institutional adjustments in the review process are necessary to truly balance the weight given to Intellectual Merit and Broader Impacts.

As such, these findings emphasize the importance of fostering a research environment in both individual projects, labs, and entire universities where Broader Impacts is not only valued in theory but also empowered in practice. This will require the NSF and supporting institutions to revisit and strengthen the mechanisms for integrating both criteria into research evaluation and career processes. The concept of an 'Impact Identity' may be one way to foster long-term research impact work that can not only benefit a researcher's career, but also drive societal good in the 'doing'. Moreover, the involvement of researchers in the review process and the integration of Broader Impacts into graduate education could facilitate earlier and more meaningful engagement with societal impacts, helping to normalize Broader Impacts as an integral part of the research process.

This increased involvement and integration may be accomplished more effectively through the support of Broader Impacts focused networks, and institutionally supported Broader Impacts centers. However, comments from interviewees indicate hesitance around such projects. It is recommended that institutionalized Broader Impacts efforts should be cautious in their approach so as not to encourage siloing or distance researchers from Broader Impacts work in the pursuit of fundability or effectiveness without regard for researcher involvement or development.

Finally, it is worth recognizing that this dissertation, written in 2024, does not fully reflect the pending changes to the Merit Review process on the horizon as a result of the 2023-2024 NSB-NSF Task Force on Merit Review. The incremental nature of changes within the NSF's Merit Review policies that are expected are still relevant, and while progress may be slow, the ongoing evolution of these criteria

offers valuable insights into the future of scientific funding in the United States. As the NSF continues to balance its internal culture with the evolving needs of the research community, the findings of this dissertation attempt to provide a snapshot of the current state of Merit Review criteria. Future work is needed to understand how the National Science Foundation's Merit Review criteria, and research impact in the United States more broadly, can develop going forward. Particularly absent from this research are the perspectives of community and industry partners who are often the collaborators, or recipients of Broader Impacts efforts. Their perspectives on and understandings of Broader Impacts have been vastly underrepresented in discussions of the Merit Review process. Given the nearly 30-year-old criterion's ultimate goal of societal benefit, the perspective, needs, and expectations of 'society' as variously defined must be considered, now more than ever.

However, it is clear that Broader Impacts, in some form, will remain a part of the research funding and award process as long as the National Science Foundation continues on its current path. Maximizing societal impact through Broader Impacts will require the National Science Foundation, and in turn the research community at large, to finally address the question: what, ultimately, is the purpose of the Broader Impacts criterion, better researchers, or better impacts - and can it someday achieve both?

8. References:

Tiffany Fleming. (n.d.). What is your impact identity? Cornell University.

National Alliance for Broader Impacts (NABI). (2018) The current state of broader impacts: advancing science and benefiting society. Columbia (MO): National Alliance for Broader Impacts. https://www.broaderimpacts.net

National Science Foundation (2023). *Proposal and Award Policies and Procedures Guide* (NSF 24-1, OMB Control Number 3145-0058). Washington, DC

NSB-NSF Staff Merit Review Task Force. (1997, March). Merit Review Final Recommendations. https://www.nsf.gov/pubs/1997/nsbmr975/nsbmr975.htm

Rothenberg, M. (2010). Making judgments about grant proposals: A brief history of the merit review criteria at the National Science Foundation. Technology & Innovation, 12(3), 189-195.

Smith, B. (2022, February 8). Building Your Research Impact Identity and Vision. https://mediaspace.msu.edu/media/Building+Your+Research+Impact+Identity+and+Vision/1_lf6tpgym

APPENDIX A

SURVEY PROTOCOL

Survey Protocol

Audience: Global Futures Scientists and Scholars Network

Topic: Broader Impacts

Screening questions:

A. How familiar are you with the term “Broader Impacts”?

   a. *Response Likert: Not at all - marginally – somewhat – reasonably - very*

B. Have you proposed, funded, or worked on any research project that fulfills or aims to fulfill a broader impacts goal, or do you expect to work on or with such a project in the future?

   a. *Response: yes – no – unsure*

C. If yes, who was the project funded by?

   a. NSF b. other (blank)

D. If yes, what was the project and/or what was the broader impacts goal in your own words?

   a. *Response: (blank)*

---

1. I have previously (or am currently) involved with or leading in the implementation of broader impacts activities or considerations in my research.
   a. *Response: Yes - no – unsure*

2. If unsure, please elaborate further:
   a. *Response: (blank)*

3. I am confident in my understanding of what the NSF Broader Impacts criterion represents.
   a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

4. Why or why not?
    a. *Response: (blank)*

5. How important to you is including broader impacts in your research?
    a. *Response Likert: not at all important - somewhat important – moderately important – very important – extremely important*

6. Why or why not?
    a. *Response: (blank)*

7. Including broader impacts in my research has enhanced the research itself.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

8. I find broader impacts requirements to be difficult to understand, vague, or imprecise.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

9. My research has benefitted from including broader impacts.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

10. My institution supports broader impacts research or activities.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

11. I have missed research opportunities because of broader impacts requirements.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

12. If possible, please elaborate.
    a. *Response: (blank)*

13. Including broader impacts considerations or activities does not enhance the value of my research.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

14. How challenging is including broader impacts activities or considerations in your research?
    a. *Response Likert: not at all challenging - somewhat challenging – moderately challenging – very challenging – extremely challenging*

15. I am interested in more training or learning opportunities around broader impacts.
    a. *Response Likert: very interested – somewhat interested – neutral – somewhat uninterested – very uninterested*

16. What kind of training would you benefit from, if any?
    a. *Response: (blank)*

17. I have proposed broader impacts activities in my research but have been unable to execute such projects for any reason.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

18. If possible, please elaborate.
    a. *Response: (blank)*

19. Including broader impacts considerations or activities takes away from time spent on my research.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

20. There are broader impacts activities I would like to incorporate into my research but have not been able to.
    a. *Response Likert: strongly agree-agree-neutral-disagree-strongly disagree*

21. If possible, please elaborate.
    a. *Response: (blank)*

22. Is there anything else you would like to share regarding your responses, or research and broader impacts in general?
    a. *Response: (blank)*

23. If you would be interested in participating in a follow-up interview about broader impacts in research for a doctoral dissertation, please provide a contact email:
    a. *Response: (blank)*

APPENDIX B

RESEARCHER SEMI-STRUCTURED INTERVIEW PROTOCOL

Researcher Interview Protocol

Verify verbal consent to recording prior to interview.

1. Can you tell me about your research focus and/or recent projects you've worked on?
2. Are there any values that drive your research activities, and can you tell me about them?
    a. You mentioned _____. Can you tell me a little more about that?
3. At a general level, what motivates you to do this work (as discussed above)?
4. How familiar would you say you are with the NSF proposal process? With the review criteria?

Artifact Elicitation Moment with Intellectual Merit Criteria:

"*The Intellectual Merit criterion encompasses the potential to advance knowledge"*

5. This definition of the Intellectual Merit Criteria comes from the NSF. Looking at this definition, can you tell me about what it means to you?
6. Personally, how do you understand intellectual merit in your own research?
    a. Do you feel confident in your understanding?
    b. How important is it to you/your work?

Artifact Elicitation Moment with Broader Impact Criteria:

"The Broader Impacts criterion encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes."

7. This definition of the Broader Impact Criteria comes from the NSF. Looking at this definition, can you tell me about what it means to you?
8. Personally, how do you understand Broader Impacts in your own research? What are some of the ways BI plays a role in your work?
    a. Do you feel confident in your understanding?
    b. How important is it to you/your work?
9. Are there any particular ways you approach balancing these two criteria in your grant applications? In your research at large?
    a. Have you faced any challenges in approaching/applying both of these criteria in your research?
    b. How have you been or how could you be further supported in proposing or doing Broader Impacts?
10. Tell me about some of your most memorable (positive) Broader Impacts that you've included in proposals.
11. What do you think the future of Broader Impacts is/should be?
    a. Are there any Broader Impacts activities or considerations you have not seen or would like to see more of?

APPENDIX C

PROGRAM OFFICER SEMI-STRUCTURED INTERVIEW PROTOCOL

Program Officer Interview Protocol

Verify verbal consent to recording prior to interview.

1. Can you tell me about yourself and your research background?
2. Can you tell me about your time as a Program Officer with the NSF?
3. What motivates you to be a PO? To do this work more broadly?

Artifact Elicitation Moment with Intellectual Merit Criteria:

"*The Intellectual Merit criterion encompasses the potential to advance knowledge"*

4. What does this definition mean to you?
    a. Are you confident in that understanding?
5. How do you think applicants understand this definition?

Artifact Elicitation Moment with Broader Impacts Criteria:

"The Broader Impacts criterion encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes."

6. What does this definition mean to you?
    a. Are you confident in that understanding?
7. How do you think applicants understand this definition?
8. How have you observed applicants approach these two criteria in grant applications and in your field at large?
9. Have you observed any
    a. Confusion or need for greater support or guidance among reviewers/applicants?
    b. Or negative sentiment or resistance towards the inclusion of the Broader Impacts criterion?
10. Are there any Broader Impacts activities or considerations you have found particularly significant and why?
11. Are there any Broader Impacts activities or considerations you have not seen or would like to see more of?
12. Is there anything else you would like to add?

APPENDIX D

MRX COMMISSION LINK LIST

Table 5. Link list for Chapter 4 (Paper 3) of collected open meeting recordings with assigned research ID, YouTube video title and date posted, and YouTube video link.

Meeting ID	YouTube Title	Date	Meeting Link
MRX 1	OPEN MRX Meeting	19-Apr-23	https://www.youtube.com/live/Er4Qv2vPBDY?si=pnALUX9E97IZj53o
MRX 2	Open MRX Meeting, May 8, 2023	8-May-23	https://youtu.be/LXoeFEPL0rk?si=AzoMTcfY1hgHmUPV
MRX 3	Open MRX Meeting	24-May-23	https://www.youtube.com/live/_b03mu-lE-A?si=1B4grNQ_zix8OFah
MRX 4	Open MRX Meeting	28-Jun-23	https://www.youtube.com/live/ulXc_cy-2bQ?si=LRS-ISiPvm-0jJz2
MRX 5	Merit Review Commission	14-Aug-23	https://www.youtube.com/live/VXGeleggyos?si=bMg85pD_eQrFVjsp
MRX 6	NSB-NSF Commission on Merit Review Open Meeting	27-Sep-23	https://www.youtube.com/live/M2e34OjPGXI?si=ywueqdzfK5p-LSM4
MRX 7	NSB-NSF Commission of Merit Review - Open Meeting	25-Oct-23	https://www.youtube.com/live/RegDeLbq9s0?si=OzgyfRdk9Kj5ghEW
MRX 8	NSB-NSF Commission on Merit Review (MRX) Meeting	28-Nov-23	https://www.youtube.com/live/VncqrySuhVw?si=F-8f7JmadsiPPrxk
MRX 9	NSB-NSF Commission on Merit Review Meeting	13-Dec-23	https://www.youtube.com/live/o0ZgGA8Xtrc?si=CQO4uv9lkDSyNmJP
MRX 10	NSB-NSF Commission on Merit Review (MRX) Meeting	18-Jan-23	https://www.youtube.com/live/XLrPe4qsPXc?si=MS9AklbY0xniW8Cb
MRX 11	NSB-NSF Commission on Merit Review (MRX) Meeting	29-Jan-24	https://www.youtube.com/live/yga6V2t4D2Y?si=B3wv9wirT-BT4U4e
MRX Plenary	NSB-NSF Commission on Merit Review - Plenary Open	10-May-24	https://youtu.be/zfXB_NDXgtI?si=1ss5Vcj942Katrvj

APPENDIX E

IRB APPROVAL

EXEMPTION GRANTED

Andrew Maynard
CGF: Future of Innovation in Society, School for the
(SFIS) 480/727-8831
Andrew.Maynard@asu.

edu Dear Andrew

Maynard:

On 10/19/2023 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Stakeholder Perspectives on Broader Impacts
Investigator:	Andrew Maynard
IRB ID:	STUDY00018855
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Stakeholder BI - Consent Form REVISED.pdf, Category: Consent Form; • Stakeholder BI - Recruitment Emails, Category: Recruitment Materials; • Stakeholder BI - Survey Consent Form REVISED.pdf, Category: Consent Form; • Stakeholders BI Research IRB REVISED.docx, Category: IRB Protocol; • Supporting Documents 09-10-2023.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2)(ii) Tests, surveys, interviews, or observation (low risk) on 10/19/2023.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are

required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB

Administrato

r cc: Dania

Wright
Dania Wright