

# A Detailed Characterization of the Expert Problem-Solving Process in Science and Engineering: Guidance for Teaching and Assessment

Argenta M. Price,<sup>†\*</sup> Candice J. Kim,<sup>‡§</sup> Eric W. Burkholder,<sup>†</sup> Amy V. Fritz,<sup>¶</sup> and Carl E. Wieman<sup>†‡</sup>

<sup>†</sup>Department of Physics, <sup>‡</sup>Graduate School of Education, <sup>§</sup>School of Medicine, and <sup>¶</sup>Department of Electrical Engineering, Stanford University, Stanford, CA 94305

## ABSTRACT

A primary goal of science and engineering (S&E) education is to produce good problem solvers, but how to best teach and measure the quality of problem solving remains unclear. The process is complex, multifaceted, and not fully characterized. Here, we present a detailed characterization of the S&E problem-solving process as a set of specific inter-linked decisions. This framework of decisions is empirically grounded and describes the entire process. To develop this, we interviewed 52 successful scientists and engineers (“experts”) spanning different disciplines, including biology and medicine. They described how they solved a typical but important problem in their work, and we analyzed the interviews in terms of decisions made. Surprisingly, we found that across all experts and fields, the solution process was framed around making a set of just 29 specific decisions. We also found that the process of making those discipline-general decisions (selecting between alternative actions) relied heavily on domain-specific predictive models that embodied the relevant disciplinary knowledge. This set of decisions provides a guide for the detailed measurement and teaching of S&E problem solving. This decision framework also provides a more specific, complete, and empirically based description of the “practices” of science.

## INTRODUCTION

Many faculty members with new graduate students and many managers with employees who are recent college graduates have had similar experiences. Their advisees/employees have just completed a program of rigorous course work, often with distinction, but they seem unable to solve the real-world problems they encounter. The supervisor struggles to figure out exactly what the problem is and how they can guide the person in overcoming it. This paper is providing a way to answer those questions in the context of science and engineering (S&E). By characterizing the problem-solving process of experts, this paper investigates the “mastery” performance level and specifies an overarching learning goal for S&E students, which can be taught and measured to improve teaching.

The importance of problem solving as an educational outcome has long been recognized, but too often postsecondary S&E graduates have serious difficulties when confronted with real-world problems (Quacquarelli Symonds, 2018). This reflects two long-standing educational problems with regard to problem solving: how to properly measure it, and how to effectively teach it. We theorize that the root of these difficulties is that good “problem solving” is a complex multifaceted process, and the details of that process have not been sufficiently characterized. Better characterization of the problem-solving process is necessary to allow problem solving, and more particularly, the complex set of skills and knowledge it entails, to be measured and taught more effectively. We sought to create an empirically grounded conceptual framework that

Jennifer Momsen, *Monitoring Editor*

Submitted Dec 2, 2020; Revised Jun 11, 2021;

Accepted Jun 23, 2021

CBE Life Sci Educ September 1, 2021 20:ar43

DOI:10.1187/cbe.20-12-0276

\*Address correspondence to: Argenta M. Price (argenta@stanford.edu).

© 2021 A. M. Price et al. CBE—Life Sciences Education © 2021 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

would characterize the detailed structure of the full problem-solving process used by skilled practitioners when solving problems as part of their work. We also wanted a framework that would allow use and comparison across S&E disciplines. To create such a framework, we examined the operational decisions (choices among alternatives that result in subsequent actions) that these practitioners make when solving problems in their discipline.

Various aspects of problem solving have been studied across multiple domains, using a variety of methods (e.g., Newell and Simon, 1972; Dunbar, 2000; National Research Council [NRC], 2012b; Lintern *et al.*, 2018). These ranged from expert self-reflections (e.g., Polya, 1945), to studies on knowledge lean tasks to discover general problem-solving heuristics (e.g., Egan and Greeno, 1974), to comparisons of expert and novice performances on simplified problems across a variety of disciplines (e.g., Chase and Simon, 1973; Chi *et al.*, 1981; Larkin and Reif, 1979; Ericsson *et al.*, 2006, 2018). These studies revealed important novice–expert differences—notably, that experts are better at identifying important features and have knowledge structures that allow them to reduce demands on working memory. Studies that specifically gave the experts unfamiliar problems in their disciplines also found that, relative to novices, they had more deliberate and reflective strategies, including more extensive planning and managing of their own behavior, and they could use their knowledge base to better define the problem (Schoenfeld, 1985; Wineburg, 1998; Singh, 2002). While these studies focused on discrete cognitive steps of the individual, an alternative framing of problem solving has been in terms of “ecological psychology” of “situativity,” looking at how the problem solver views and interacts with the environment in terms of affordances and constraints (Greeno, 1994). “Naturalistic decision making” is a related framework that specifically examines how experts make decisions in complex, real-world, settings, with an emphasis on the importance of assessing the situation surrounding the problem at hand (Klein, 2008; Mosier *et al.*, 2018).

While this work on expertise has provided important insights into the problem-solving process, its focus has been limited. Most has focused on looking for cognitive differences between experts and novices using limited and targeted tasks, such as remembering the pieces on a chessboard (Chase and Simon, 1973) or identifying the important concepts represented in an introductory physics textbook problem (Chi *et al.*, 1981). It did not attempt to explore the full process of solving, particularly for solving the type of complex problem that a scientist or engineer encounters as a member of the workforce (“authentic problems”).

There have also been many theoretical proposals as to expert problem-solving practices, but with little empirical evidence as to their completeness or accuracy (e.g., Polya, 1945; Heller and Reif, 1984; Organisation for Economic Cooperation and Development [OECD], 2019). The work of Dunbar (2000) is a notable exception to the lack of empirical work, as his group did examine how biologists solved problems in their work by analyzing lab meetings held by eight molecular biology research groups. His groundbreaking work focused on creativity and discovery in the research process, and he identified the importance of analogical reasoning and distributed reasoning by scientists in answering research questions and gaining new insights.

Kozma *et al.* (2000) studied professional chemists solving problems, but their work focused only on the use of specialized representations.

The “cognitive systems engineering” approach (Lintern *et al.*, 2018) takes a more empirically based approach looking at experts solving problems in their work, and as such tends to span aspects of both the purely cognitive and the ecological psychological theories. It uses both observations of experts in authentic work settings and retrospective interviews about how experts carried out particular work tasks. This theoretical framing and the experimental methods are similar to what we use, particularly in the “naturalistic decision making” area of research (Mosier *et al.*, 2018). That work looks at how critical decisions are made in solving specific problems in their real-world setting. The decision process is studied primarily through retrospective interviews about challenging cases faced by experts. As described below, our methods are adapted from that work (Crandall *et al.*, 2006), though there are some notable differences in focus and field. A particular difference is that we focused on identifying *what are* decisions to be made, which are more straight-forward to identify from retrospective interviews than *how* those decisions are made. We all have the same ultimate goal, however, to improve the training/teaching of the respective expertise.

Problem solving is central to the processes of science, engineering, and medicine, so research and educational standards about scientific thinking and the process and practices of science are also relevant to this discussion. Work by Osborne and colleagues describes six styles of scientific reasoning that can be used to explain how scientists and students approach different problems (Kind and Osborne, 2016). There are also numerous educational standards and frameworks that, based on theory, lay out the skills or practices that science and engineering students are expected to master (e.g., American Association for the Advancement of Science [AAAS], 2011; Next Generation Science Standards Lead States, 2013; OECD, 2019; ABET, 2020). More specifically related to the training of problem solving, Priemer *et al.* (2020) synthesizes literature on problem solving and scientific reasoning to create a “STEM [science, technology, engineering, and mathematics] and computer science framework for problem solving” that lays out steps that could be involved in a students’ problem-solving efforts across STEM fields. These frameworks provide a rich groundwork, but they have several limitations: 1) They are based on theoretical ideas of the practice of science, not empirical evidence, so while each framework contains overlapping elements of the problem-solving process, it is unclear whether they capture the complete process. 2) They are focused on school science, rather than the actual problem solving that practitioners carry out and that students will need to carry out in future STEM careers. 3) They are typically underspecified, so that the steps or practices apply generally, but it is difficult to translate them into measurable learning goals for students to practice. Working to address that, Clemmons *et al.* (2020) recently sought to operationalize the core competencies from the *Vision and Change* report (AAAS, 2011), establishing a set of skills that biology students should be able to master.

Our work seeks to augment this prior work by building a conceptual framework that is empirically based, grounded in how scientists and engineers solve problems in practice instead

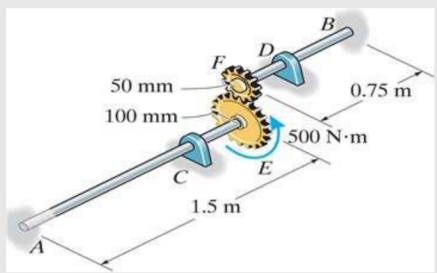
<p><b>Mechanical Engineering</b> “The two shafts shown in the figure below are made of A-36 steel. Each has a diameter of 25 mm and they are connected using the gears fixed to their ends. Their other ends are attached to fixed supports at A and B. They are also supported by journal bearings at C and D, which allow free rotation of the shafts along their axes. If a torque of 500 N.m is applied to the gear at E as shown in the figure, determine the reactions at A &amp; B. For A-36 Steel: <math>E = 200</math> GPa and <math>G = 75</math> GPa.”</p>	
<p><b>Physics</b> “Solve for the eigenvalues and eigenfunctions of the hydrogen atom using the Dirac equation.”</p>	
<p><b>Biology</b> “Comparisons of the patterns of mRNA levels across different human cell types show that the level of expression of almost every active gene is different. The patterns of mRNA abundance are so characteristic of cell type that they can be used to determine the tissue of origin of cancer cells, even though the cancer cells may have metastasized to different parts of the body. By definition, however, cancer cells are different from their noncancerous precursor cells. How do you suppose then that patterns of mRNA expression might be used to determine the tissue source of a human cancer?”</p>	

FIGURE 1. Example problems from courses or textbooks in mechanical engineering, physics and biology. Problems from: Mechanical engineering: Wayne State mechanical engineering sample exam problems (Wayne State, n.d.), Physics: A standard physics problem in nearly every advanced quantum mechanics course, Biology: Molecular Biology of the Cell 6th edition, Chapter 7 end of chapter problems (Alberts et al., 2014).

of in school. We base our framework on the decisions that need to be made during problem solving, which makes each item clearly defined for practice and assessment. In our analysis of expert problem solving, we empirically identified the entire problem-solving process. We found this includes deciding when and how to use the steps and skills defined in the work described previously but also includes additional elements. There are also questions in the literature about how generalizable across fields a particular set of practices may be. Here, we present the first empirical examination of the entire problem-solving process, and we compare that process across many different S&E disciplines.

A variety of instructional methods have been used to try and teach science and engineering problem solving, but there has been little evidence of their efficacy at improving problem solving (for a review, see NRC, 2012b). Research explicitly on teaching problem solving has primarily focused on textbook-type exercises and utilized step-by-step strategies or heuristics. These studies have shown limited success, often getting students to follow specific procedural steps but with little gain in actually solving problems and showing some potential drawbacks (Heller and Reif, 1984; Heller et al., 1992; Huffman, 1997; Heckler, 2010; Kuo et al., 2017). As discussed later, the framework presented here offers guidance for different and potentially more effective approaches to teaching problem solving.

These challenges can be illustrated by considering three different problems taken from courses in mechanical engineering, physics, and biology, respectively (Figure 1). All of these problems are challenging, requiring considerable knowledge and effort by the student to solve correctly. Problems such as these are routinely used to both assess students' problem-solving

skills, and students are expected to learn such skills by practicing doing such problems. However, it is obvious to any expert in the respective fields, that, while these problems might be complicated and difficult to answer, they are vastly different from solving authentic problems in that field. They all have well-defined answers that can be reached by straightforward solution paths. More specifically, they do not involve needing to use judgment to make any decisions based on limited information (e.g., insufficient to specify a correct decision with certainty). The relevant concepts and information and assumptions are all stated or obvious. The failure of problems like these to capture the complexity of authentic problem solving underlies the failure of efforts to measure and teach problem solving. Recognizing this failure motivated our efforts to more completely characterize the problem-solving process of practicing scientists, engineers, and doctors.

We are building on the previous work studying expert–novice differences and problem solving but taking a different direction. We sought to create an empirically grounded framework that would characterize the detailed structure of the full problem-solving process by focusing on the operational decisions that skilled practitioners make when successfully solving authentic problems in their scientific, engineering, or medical work. We chose to identify the decisions that S&E practitioners made, because, unlike potentially nebulous skills or general problem-solving steps that might change with the discipline, decisions are sufficiently specified that they can be individually practiced by students and measured by instructors or departments. The authentic problems that we analyzed are typical problems practitioners encounter in “doing” the science or engineering entailed in their jobs. In the language of traditional problem-solving and expertise research, such authentic problems are

“ill-structured” (Simon, 1973) and require “adaptive expertise” (Hatano and Inagaki, 1986) to solve. However, our authentic problems are considerably more complex and unstructured than what is normally considered in those literatures, because not only do they lack a clear solution path, but in many cases, it is not clear a priori that they have any solution at all. Determining that, and whether the problem needs to be redefined to be solvable, is part of the successful expert solution process. Another way in which our set of decisions goes beyond the characterization of what is involved in adaptive expertise is the prominent role of making judgments with limited information.

A common reaction of scientists and engineers to seeing the list of decisions we obtain as our primary result is, “Oh, yes, these are things I always do in solving problems. There is nothing new here.” It is comforting that these decisions all look familiar; that supports their validity. However, what is new is not that experts are making such decisions, but rather that there is a relatively small but complete set of decisions that has now been explicitly identified and that applies so generally.

We have used a much larger and broader sample of experts in this work than used in prior expert–novice studies, and we used a more stringent selection criterion. Previous empirical work has typically involved just a few experts, almost always in a single domain, and included graduate students as “experts” in some cases. Our semistructured interview sample was 31 experienced practitioners from 10 different disciplines of science, engineering, and medicine, with demonstrated competence and accomplishments well beyond those of most graduate students. Also, approximately 25 additional experts from across science, engineering, and medicine served as consultants during the planning and execution of this work.

Our research question was: What are the decisions experts make in solving authentic problems, and to what extent is this set of decisions to be made consistent both within and across disciplines?

Our approach was designed to identify the level of consistency and unique differences across disciplines. Our hypothesis was that there would be a manageable number (20–50) of decisions to be made, with a large amount of overlap of decisions made between experts within each discipline and a substantial but smaller overlap across disciplines. We believed that if we had found that every expert and/or discipline used a large and completely unique set of decisions, it would have been an interesting research result but of little further use. If our hypothesis turned out to be correct, we expected that the set of decisions obtained would have useful applications in guiding teaching and assessment, as they would show how experts in the respective disciplines applied their content knowledge to solve problems and hence provide a model for what to teach. We were not expecting to find the nearly complete degree of overlap in the decisions made across all the experts.

## METHODS

We first conducted 22 relatively unstructured interviews with a range of S&E experts, in which we asked about problem-solving expertise in their fields. From these interviews, we developed an initial list of decisions to be made in S&E problem solving. To refine and validate the list, we then carried out a set of 31 semistructured interviews in which S&E experts chose a specific problem from their work and described the solution process in

detail. The semistructured interviews were coded for the decisions represented, either explicitly stated or implied by a choice of action. This provided a framework of decisions that characterize the problem-solving process across S&E disciplines. The research was approved by the Stanford Institutional Review Board (IRB no. 48785), and informed consent was obtained from all the participants.

This work involved interviewing many experts across different fields. We defined experts as practicing scientists, engineers, or physicians with considerable experience working as faculty at highly rated universities or having several years of experience working in moderately high-level technical positions at successful companies. We also included a few longtime postdocs and research staff in biosciences to capture more details of experimental decisions from which faculty members in those fields often were more removed. This definition of expert allows us to identify the practices of skilled professionals; we are not studying what makes only the most exceptional experts unique.

Experts were volunteers recruited through direct contact via the research team's personal and professional networks and referrals from experts in our networks. This recruitment method likely biased our sample toward people who experienced relatively similar training (most were trained in STEM disciplines at U.S. universities within the last 15–50 years). Within this limitation, we attempted to get a large range of experts by field and experience. This included people from 10 different fields (including molecular biology/biochemistry, ecology, and medicine), 11 U.S. universities, and nine different companies or government labs, and the sample was 33% female (though our engineering sample only included one female). The medical experts were volunteers from a select group of medical school faculty chosen to serve as clinical reasoning mentors for medical students at a prestigious university. We only contacted people who met our criteria for being an “expert,” and everyone who volunteered was included in the study. Most of the people who were contacted volunteered, and the only reason given for not volunteering was insufficient time. Other than their disciplinary expertise, there was little to distinguish these experts beyond the fact they were acquaintances with members of the team or acquaintances of acquaintances of team or project advisory board members. The precise number from each field was determined largely by availability of suitable experts.

We defined an “authentic problem” to be one that these experts solve in their actual jobs. Generally, this meant research projects for the science and engineering faculty, design problems for the industry engineers, and patient diagnoses for the medical doctors. Such problems are characterized by complexity, with many factors involved and no obvious solution process, and involve substantial time, effort, and resources. Such problems involve far more complexity and many more decisions, particularly decisions with limited information, than the typical problems used in previous problem-solving research or used with students in instructional settings.

### Creating an Initial List of Problem-Solving Decisions

We first interviewed 22 experts (Table 1), most of whom were faculty at a prestigious university, in which we asked them to discuss expertise and problem solving in their fields as it related to their own experiences. This usually resulted in their discussing examples of one or more problems they had solved. Based



on the first seven interviews, plus reflections on personal experience from the research team and review of the literature on expert problem solving and teaching of scientific practices (Ericsson *et al.*, 2006; NRC, 2012a; Wieman, 2015), we created a generic list of decisions that were made in S&E problem solving. In the rest of the unstructured interviews (15), we also provided the experts with our list and asked them to comment on any additions or deletions they would suggest. Faculty who had close supervision of graduate students and industry experts who had extensively supervised inexperienced staff were particularly informative. Their observations of the way inexperienced people could fail made them sensitive to the different elements of expertise and where incorrect decisions could be made. Although we initially expected to find substantial differences across disciplines, from early in the process, we noted a high degree of overlap across the interviews in the decisions that were described.

### Refinement and Validation of the List of Decisions

After creating the preliminary list of decisions from the informal interviews, we conducted a separate set of more structured interviews to test and refine the list. Semistructured interviews were conducted with 31 experts from across science, engineering, and medical fields (Table 1). For these interviews, we recruited experts from a range of universities and companies, though the range of institutions is still limited, given the sample size. Interviews were conducted in person or over video chat and were transcribed for analysis. In the semistructured interviews, experts were asked to choose a problem or two from their work that they could recall the details of solving and then describe the process, including all the steps and decisions they made. So that we could get a full picture of the successful problem-solving process, we decided to focus the interviews on problems that they had eventually solved successfully, though their processes inherently involved paths that needed to be revised and reconsidered. Transcripts from interviewees who agreed to have their interview transcript published are available in the supplemental data set.

Our interview protocol (see Supplemental Text) was inspired in part by the critical decision method of cognitive task analysis (Crandall *et al.*, 2006; Lintern *et al.*, 2018), which was created for research in cognitive systems engineering and naturalistic decision making. There are some notable differences between our work and theirs, both in research goal and method. First, their goal is to improve training in specific fields by focusing on *how* critical decisions are made in that field during an unusual or important event; the analysis seeks to identify factors involved in making those critical decisions. We are focusing on the overall problem solving and how it compares across many different fields, which quickly led to attention on *what* decisions are to be made, rather than how a limited set of those decisions are made. We asked experts to describe a specific, but not necessarily unusual, problem in their work, and focused our analysis on identifying all decisions made, not reasons for making them or identifying which were most critical. The specific order of problem-solving steps was also less important to us, in part because it was clear that there was no consistent order that was followed. Second, we are looking at different types of work. Cognitive systems engineering work has primarily focused on performance in professions like firefighters, power plant opera-

tors, military technicians, and nurses. These tend to require time-sensitive critical skills that are taught with modest amounts of formal training. We are studying scientists, engineers, and doctors solving problems that require much longer and less time-critical solutions and for which the formal training occupies many years.

Given our different focus, we made several adaptations to eliminate some of the more time-consuming steps from the interview protocol, allowing us to limit the interview time to approximately 1 hour. Both protocols seek to elicit an accurate and complete reporting of the steps taken and decisions made in the process of solving a problem. Our general strategy was: 1) Have the expert explain the problem and talk step by step through the decisions involved in solving it, with relatively few interruptions from the interviewer except to keep the discussion focused on the specific problem and occasionally to ask for clarifications. 2) Ask follow-up questions to probe for more detail about particular steps and aspects of the problem-solving process. 3) Occasionally ask for general thoughts on how a novice's process might differ.

While some have questioned the reliability of information from retrospective interviews (Nisbett and Wilson, 1977), we believe we avoid these concerns, because we are only identifying a decision to be made, which in this case, means identifying a well-defined action that was chosen from alternatives. This is less subjective and much more likely to be accurately recalled than is the rationale behind such a decision. See Ericsson and Simon (1980). However, the decisions identified may still be somewhat limited—the process of deciding among possible actions might involve additional decisions in the moment, when the solution is still unknown, that we are unable to capture in the retrospective context. For the decisions we can identify, we are able to check their accuracy and completeness by comparing them with the actions taken in the conduct of the research/design. For example, consider this quote from a physician who had to re-evaluate a diagnosis, “And, in my very subjective sense, he seemed like he was being forthcoming and honest. Granted people can fool you, but he seemed like he was being forthcoming. So we had to reevaluate.” The physician then considered alternative diagnoses that could explain a test result that at first had indicated an incorrect diagnosis. While this quote does describe the (retrospective) reasoning behind a decision, we do not need to know whether that reasoning is accurately recalled. We can simply code this as “decision 18, how believable is info?” The physician followed up by considering alternative diagnoses, which in this context was coded as “26, how good is solution?” and “8, potential solutions?” This was followed by the description of the literature and additional tests conducted. These indicated actions taken that confirm the physician made a decision about the reliability of the information given by the patient.

### Interview Coding

We coded the semistructured interviews in terms of decisions made, through iterative rounds of coding (Chi, 1997), following a “directed content analysis approach,” which involves coding according to predefined theoretical categories and updating the codes as needed based on the data (Hsieh and Shannon, 2005). Our predefined categories were the list of decisions we had developed during the informal interviews. This approach means that we limited the focus of our qualitative analysis—we

were able to test and refine the list of decisions, but we did not seek to identify all possible categories of approach to selecting and solving problems. The goals of each iterative round of coding are described in the next three paragraphs. To code for decisions in general, we matched decisions from the list to statements in each interview, based on the following criteria: 1) there was an explicit statement of a decision or choice made or needing to be made; 2) there was the description of the outcome of a decision, such as listing important features of the problem (that had been decided on) or conclusions arrived at; or 3) there was a statement of actions taken that indicated a decision about the appropriate action had been made, usually from a set of alternatives. Two examples illustrate the types of comments we identified as decisions: A molecular biologist explicitly stated the decisions required to decompose a problem into subproblems (decision 11), “Which cell do we use? The gene. Which gene do we edit? Which part of that gene do we edit? How do we build the enzyme that is going to do the cutting? ... And how do we read out that it worked?” An ecologist made a statement that was also coded as a decomposition decision, because it described the action taken: “So I analyze the bird data first on its own, rather than trying to smash all the taxonomic groups together because they seem really apples and oranges. And just did two kinds of analysis, one was just sort of across all of these cases, around the world.” A single statement could be coded as multiple decisions if they were occurring simultaneously in the story being recalled or were intimately interconnected in the context of that interview, as with the ecology quote, in which the last sentence leads into deciding what data analysis is needed. Inherent in nearly every one of these decisions was that there was insufficient information to know the answer with certainty, so judgment was required.

Our primary goal for the first iterative round of coding was to check whether our list was complete by checking for any decisions that were missing, as indicated by either an action taken or a stated decision that was not clearly connected to a decision on our initial list. In this round, we also clarified wording and combined decisions that we were consistently unable to differentiate during the coding. A sample of three interviews (from biology, medicine, and electrical engineering) were first coded independently by four coders (AP, EB, CK, and AF), then discussed. The decision list was modified to add decisions and update wording based on that discussion. Then the interviews were recoded with the new list and rediscussed, leading to more refinements to the list. Two additional interviews (from physics and chemical engineering) were then coded by three coders (AP, EB, and CK) and further similar refinements were made. Throughout the subsequent rounds of coding, we continued to check for missing decisions, but after the additions and adjustments made based on these five interviews, we did not identify any more missing decisions.

In our next round of coding, we focused on condensing overlapping decisions and refining wording to improve the clarity of descriptions as they applied across different disciplinary contexts and to ensure consistent interpretation by different coders. Two or three coders independently coded an additional 11 interviews, iteratively meeting to discuss codes identified in the interviews, refining wording and condensing the list to improve agreement and combine overlapping codes, and then using the updated list to code subsequent interviews. We condensed the

list by combining decisions that represented the same cognitive process taking place at different times, that were discipline-specific variations on the same decision, or that were substeps involved in making a larger decision. We noticed that some decisions were frequently co-coded with others, particularly in some disciplines. But if they were identified as distinct a reasonable fraction of the time in any discipline, we listed them as separate. This provided us with a list, condensed from 42 to 29 discrete decisions (plus five additional non-decision themes that were so prevalent that they are important to describe), that gave good consistency between coders.

Finally, we used the resulting codes to tabulate which decisions occurred in each interview, simplifying our coding process to focus on deciding whether or not each decision had occurred, with an example if it did occur to back up the “yes” code, but no longer attempting to capture every time each decision was mentioned. Individual coders identified decisions mentioned in the remaining 15 interviews. Interviews that had been coded with the early versions of the list were also recoded to ensure consistency. Coders flagged any decisions they were unsure about occurring in a particular interview, and two to four coders (AP, EB, CK, and CW) met to discuss those debated codes, with most uncertainties being resolved by explanations from a team member who had more technical expertise in the field of the interview. Minor wording changes were made during this process to ensure that each description of a decision captured all instantiations of the decision across disciplines, but no significant changes to the list were needed or made.

Coding an interview in terms of decisions made and actions taken in the research often required a high level of expertise in the discipline in question. The coder had to be familiar with the conduct of research in the field in order to recognize which actions corresponded to a decision between alternatives, but our team was assembled with this requirement in mind. It included high-level expertise across five different fields of science, engineering, and medicine and substantial familiarity with several other fields.

Supplemental Table S1 shows the final tabulation of decisions identified in each interview. In the tabulation, most decisions were marked as either “yes” or “no” for each interview, though 65 out of 1054 total were marked as “implied,” for one of the following reasons: 1) for 40/65, based on the coder’s knowledge of the field, it was clear that a step must have been taken to achieve an outcome or action, even though that decision was not explicitly mentioned (e.g., interviewees describe collecting certain raw data and then coming to a specific conclusion, so they must have decided how to analyze the data, even if they did not mention the analysis explicitly); 2) for 15/65, the interview context was important, in that multiple statements from different parts of the interview taken together were sufficient to conclude that the decision must have happened, though no single statement described that decision explicitly; 3) 10/65 involved a decision that was explicitly discussed as an important step in problem solving, but they did not directly state how it was related to the problem at hand, or it was stated only in response to a direct prompt from the interviewer. The proportion of decisions identified in each interview, broken down by either explicit or explicit + implied, is presented in Supplemental Tables S1 and S2. Table 2 and Figure 2 of the main text show explicit + implied decision numbers.

TABLE 1. Number of interviews conducted, by field of interviewee

Discipline	Informal interviews (creation of initial list)	Structured interviews (validation/refinement)	Notes
Biology (5 biochem/molecular bio, 2 cell bio, 1 plant bio, 1 immunology, 1 ecology)	2	8	Female: 6, URM: 2 5 faculty, 2 industry 3 acad staff/postdoc (year 5+)
Medicine (6 internal med or pediatrics, 1 oncology, 2 surgery)	4	6	Female: 4, URM: 1 All medical faculty
Physics (4 experiment, 3 theory)	2	5	Female: 1, URM: 1 All faculty
Electrical Engineering	4	3	2 faculty, 4 industry, 1 acad. staff
Chemical Engineering	2	2	Female: 1 3 industry, 1 acad. staff
Mechanical Engineering	2	2	URM: 1, 2 faculty, 2 industry
Earth Science	1	2	Female: 2, 2 faculty, 1 industry
Chemistry	1	2	Female: 2, all faculty
Computer Science	2	1	Female: 1, 2 faculty, 1 industry
Biological Engineering	2	–	All faculty or acad. staff
Total	22	31	Female: 17, URM: 5

URM (under-represented minority) included 3 African American and 2 Hispanic/Latinx. One medical faculty member was interviewed twice – in both informal and structure interviews, for a total of 53 interviews with 52 experts.

Two of the interviews that had not been discussed during earlier rounds of coding (one physics [AP and EB], one medicine [AP and CK]) were independently coded by two coders to check interrater reliability using the final list of decisions. The goal of our final coding was to tabulate whether or not each expert described making each decision at any point in the problem-solving process, so the level of detail we chose for coding and interrater reliability was whether or not a decision was present in the entire interview. The decisions identified in each interview were compared for the two coders. For both interviews, the raters disagreed on whether or not only one of the 29 decisions occurred. Codes of “implied” were counted as agreement if the other coder selected either “yes” or “implied.” This equates to a percent agreement of 97% for each interview (28 agree/29 total decisions per interview = 97%). As a side note, there was also one disagreement per interview on the coding of the five other themes, but those themes were not a focus of this work nor the interviews.

## RESULTS

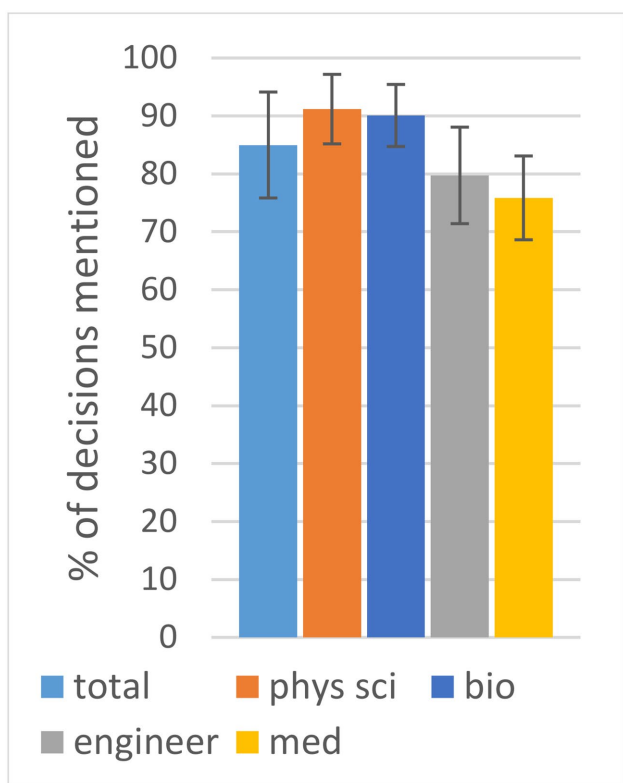
We identified a total set of 29 decisions to be made (plus five other themes), all of which were identified in a large fraction of the interviews across all disciplines (Table 2 and Figure 2). There was a surprising degree of overlap across the different fields with all the experts mentioning similar decisions to be made. All 29 were evident by the fifth semistructured interview, and on average, each interview revealed 85% of the 29 decisions. Many decisions occurred multiple times in an interview, with the number of times varying widely, depending on the length and complexity of the problem-solving process discussed.

We focused our analysis on what decisions needed to be made, not on the experts’ processes for making those decisions: noting that a choice happened, not *how* they selected and chose among different alternatives. This is because, while the decisions to be made were the same across disciplines, how the experts made those decisions varied greatly by discipline and individual. The process of making the decisions relied on

specialized disciplinary knowledge and experience and may vary depending on demographics or other factors that our study design (both our sample and nature of retrospective interviews) did not allow us to investigate. However, while that knowledge was distinct and specialized, we could tell that it was consistently organized according to a common structure we call a “predictive framework,” as discussed in the “*Predictive Framework*” section below. Also, while every “decision” reflected a step in the problem solving involved in the work, and the expert being interviewed was involved in making or approving the decision, that does not mean the decision process was carried out only by that individual. In many cases, the experts described the decisions made in terms of ideas and results of their teams, and the importance of interpersonal skills and teamwork was an important non-decision theme raised in all interviews.

We were particularly concerned with the correctness and completeness of the set of decisions. Although the correctness was largely established by the statements in the interviews, we also showed the list of decisions to these experts at the end of the interviews as well as to about a dozen other experts. In all cases, they all agreed that these decisions were ones they and others in their field made when solving problems. The completeness of the list of decisions was confirmed by: 1) looking carefully at all specific actions taken in the described problem-solving process and checking that each action matched a corresponding decision from the list; and 2) the high degree of consistency in the set of decisions across all the interviews and disciplines. This implies that it is unlikely that there are important decisions that we are missing, because that would require any such missing decisions to be consistently unspoken by all 31 interviewees as well as consistently unrecognized by us from the actions that were taken in the problem-solving process.

In focusing on experts’ recollections of their *successful* solving of problems, our study design may have missed decisions that experts only made during failed problem-solving attempts. However, almost all interviews described solution paths that were not smooth and continuous, but rather involved going



**FIGURE 2.** Proportion of decisions coded in interviews by field. This tabulation includes decisions 1–29, not the additional themes. Error bars represent standard deviations. Number of interviews: total = 31; physical science = 9; biological science = 8; engineering = 8; medicine = 6. Compared with the sciences, slightly fewer decisions overall were identified in the coding of engineering and medicine interviews, largely for discipline-specific reasons. See Supplemental Table S2 and associated discussion.

down numerous dead ends. There were approaches that were tried and failed, data that turned out to be ambiguous and worthless, and so on. Identifying the failed path involved reflection decisions (23–26). Often decision 9 (is problem solvable?) would be mentioned, because it described a path that was determined to be not solvable. For example, a biologist explained, “And then I ended up just switching to a different strain that did it [crawling off the plate] less. Because it was just ... hard to really get them to behave themselves. I suppose if I really needed to rely on that very particular one, I probably would have exhausted the possibilities a bit more.” Thus, we expect unsuccessful problem solving would entail a smaller subset of decisions being made, particularly lack of reflection decisions, or poor choices on the decisions, rather than making a different set of decisions.

The set of decisions represent a remarkably consistent structure underlying S&E problem solving. For the purposes of presentation, we have categorized the decisions as shown in Figure 3, roughly based on the purposes they achieve. However, the process is far less orderly and sequential than implied by this diagram, or in fact any characterization of an orderly “scientific method.” We were struck by how variable the sequence of decisions was in the descriptions provided. For example, experts who described how they began work on a

problem sometimes discussed importance and goals (1–3, what is important in field?; opportunity fits solver’s expertise?; and goals, criteria, constraints?), but others mentioned a curious observation (20, any significant anomalies?), important features of their system that led them to questions (4, important features and info?, 6, how to narrow down problem?), or other starting points. We also saw that there were flexible connections between decisions and repeated iterations—jumping back to the same type of decision multiple times in the solution process, often prompted by reflection as new information and insights were developed. The sequence and number of iterations described varied dramatically by interview, and we cannot determine to what extent this was due to legitimate differences in the problem-solving process or to how the expert recalled and chose to describe the process. This lack of a consistent starting point, with jumping and iterating between decisions, has also been identified in the naturalistic decision-making literature (Mosier *et al.*, 2018). Finally, the experts also often described considering multiple decisions simultaneously. In some interviews, a few decisions were always described together, while in others, they were clearly separate decisions. In summary, while the specific decisions themselves are fully grounded in expert practice, the categories and order shown here are artificial simplifications for presentation purposes.

The decisions contained in the seven categories are summarized here. See Supplemental Table S2 for specific examples of each decision across multiple disciplines.

### Category A. Selection and Goals of the Problem

This category involves deciding on the importance of the problem, what criteria a solution must meet, and how well it matches the capabilities, resources, and priorities of the expert. As an example, an earth scientist described the goal of her project (decision 3, goals, criteria, constraints?) to map and date the earliest volcanic rocks associated with what is now Yellowstone and explained why the project was a good fit for her group (2, opportunity fits solver’s expertise?) and her decision to pursue the project in light of the significance of this type of eruption in major extinction events (1, what is important in field?). In many cases, decisions related to framing (see category B) were mentioned before decisions in this category or were an integral part of the process for developing goals.

Decisions in this category are:

1. **What is important in the field?**  
What are important questions or problems? Where is the field heading? Are there advances in the field that open new possibilities?
2. **Opportunity fits solver’s expertise?**  
If and where are there gaps/opportunities to solve in field? Given experts’ unique perspectives and capabilities, are there opportunities particularly accessible to them? (This could involve challenging the status quo, questioning assumptions in the field.)
3. **Goals, criteria, constraints?**  
What are the goals for this problem? Possible considerations include:
  - a. What are the goals, design criteria, or requirements of the problem or its solution?



TABLE 2. Problem-solving decisions and percentages of expert interviews in which they occur<sup>a</sup>

A. Selection and goals (Occur in 100% <sup>b</sup> )	B. Frame problem (100%)	C. Plan process for solving (100%)	D. Interpret info and choose solutions (100%)	E. Reflect <sup>c</sup> (100%)	F. Implications and communicate results (84%)
1. <sup>c</sup> (61%) What is important in field?	4. (100%) Important features and info?	10. (100%) Approximations and simplifications to make?	16. (81%) Which calculations and data analysis?	23. (77%) Assumptions and simplifications appropriate?	27. (65%) Broader implications?
2. (77%) Opportunity fits solver's expertise?	5. (100%) What predictive framework? <sup>d</sup>	11. (68%) How to decompose into sub-problems?	17. (68%) How to represent and organize information?	24. (84%) Additional knowledge needed?	28. (55%) Audience for communication?
3. (100%) Goals, criteria, constraints?	6. (97%) How to narrow down problem?	12. (90%) Most difficult or uncertain areas?	18. (77%) How believable is information?	25. (94%) How well is solving approach working?	29. (68%) Best way to present work?
	7. (97%) Related problems?	13. (100%) What info needed?	19. (100%) How does info compare to predictions?	26. (100%) How good is solution?	
	8. (100%) Potential Solutions?	14. (87%) Priorities?	20. (71%) Any significant anomalies?		
	9. (74%) Is problem solvable?	15. (100%) Specific plan for getting information?	21. (97%) Appropriate conclusions?		
			22. (97%) What is best solution?		

<sup>a</sup>See supplementary text and Table S2 for full description and examples of each decision. A set of other non-decision knowledge and skill development themes were also frequently mentioned as important to professional success: Staying up to date in the field (84%), intuition and experience (77%), interpersonal and teamwork (100%), efficiency (32%), and attitude (68%).

<sup>b</sup>Percentage of interviews in which category or decision was mentioned.

<sup>c</sup>Numbering is for reference. In practice ordering is fluid – involves extensive iteration with other possible starting points.

<sup>d</sup>Chosen predictive framework(s) will inform all other decisions.

<sup>e</sup>Reflection occurs throughout process, and often leads to iteration. Reflection on solution occurs at the end as well.

- b. What is the scope of the problem?
- c. What constraints are there on the solution?
- d. What will be the criteria on which the solution is evaluated?

**Category B. Frame Problem**

These decisions lead to a more concrete formulation of the solution process and potential solutions. This involves identifying the key features of the problem and deciding on predictive frameworks to use (see “Predictive Framework” section below), as well as narrowing down the problem, often forming specific questions or hypotheses. Many of these decisions are guided by past problem solutions with which the expert is familiar and sees as relevant. The framing decisions of a physician can be seen in his discussion of a patient with liver failure who had previously been diagnosed with HIV but had features (4, important features and info?; 5, what predictive framework?) that made the physician question the HIV diagnosis (5, what predictive framework?; 26, how good is solution?). His team then searched for possible diagnoses that could explain liver failure and lead to a false-positive HIV test (7, related problems?; 8, potential solutions?), which led to their hypothesis the patient might have Q fever (6, how to narrow down problem?; 13, what info needed?;

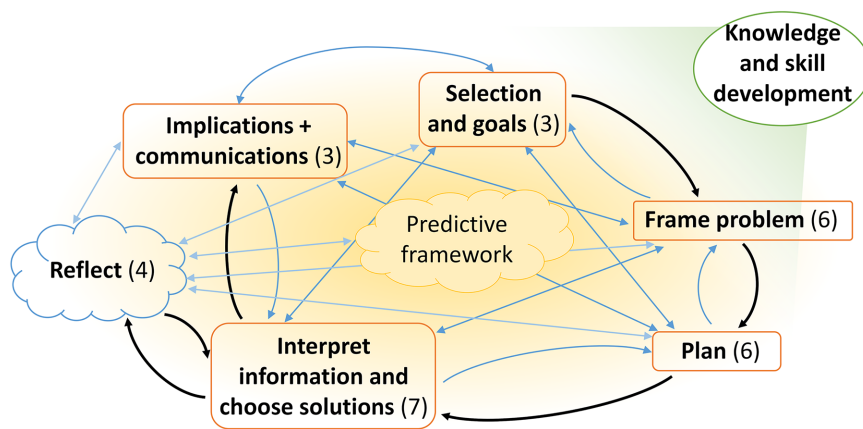


FIGURE 3. Representation of problem-solving decisions by categories. The black arrows represent a hypothetical but unrealistic order of operations, the blue arrows represent more realistic iteration paths. The decisions are grouped into categories for presentation purposes; numbers indicate the number of decisions in each category. Knowledge and skill development were commonly mentioned themes but are not decisions.

15, specific plan for getting info?). While each individual decision is strongly supported by the data, the categories are groupings for presentation purposes. In particular, framing (category B) and planning (see category C) decisions often blended together in interviews.

Decisions in this category are:

4. **Important features and info?**  
What are the important underlying features or concepts that apply? Could include:
  - a. Which available information is relevant to problem solving and why?
  - b. (When appropriate) Create/find a suitable abstract representation of core ideas and information Examples: physics, equation representing process involved; chemistry, bond diagrams/potential energy surfaces; biology, diagram of pathway steps.
5. **What predictive framework?**  
Which potential predictive frameworks to use? (Decide among possible predictive frameworks or create framework.) This includes deciding on the appropriate level of mechanism and structure that the framework needs to embody to be most useful for the problem at hand.
6. **How to narrow down the problem?**  
How to narrow down the problem? Often involves formulating specific questions and hypotheses.
7. **Related problems?**  
What are related problems or work seen before, and what aspects of their problem-solving process and solutions might be useful in the present context? (This may involve reviewing literature and/or reflecting on experience.)
8. **Potential solutions?**  
What are potential solutions? (This is based on experience and fitting some criteria for solution they have for a problem having general key features identified.)
9. **Is problem solvable?**  
Is the problem plausibly solvable and is the solution worth pursuing given the difficulties, constraints, risks, and uncertainties?

### Category C. Plan the Process for Solving

These decisions establish the specifics needed to solve the problem and include: how to simplify the problem and decompose it into pieces, what specific information is needed, how to obtain that information, and what are the resources needed and priorities? Planning by an ecologist can be seen in her extensive discussion of her process of simplifying (10, approximations/simplifications to make?) a meta-analysis project about changes in migration behavior, which included deciding what types of data she needed (13, what info needed?), planning how to conduct her literature search (15, specific plan for getting info?), difficulties in analyzing the data (12, most difficult/uncertain areas?; 16, which calculations and data analysis?), and deciding to analyze different taxonomic groups separately (11, how to decompose into subproblems?). In general, decomposition often resulted in multiple iterations through the problem-solving decisions, as subsets of decisions need to be made about each decomposed aspect of a problem. Framing (category B) and

planning (category C) decisions occupied much of the interviews, indicating their importance.

Decisions in this category are:

10. **Approximations and simplifications to make?**  
What approximations or simplifications are appropriate? How to simplify the problem to make it easier to solve? Test possible simplifications/approximations against established criteria.
11. **How to decompose into subproblems?**  
How to decompose the problem into more tractable subproblems? (Subproblems are independently solvable pieces with their own subgoals.)
12. **Most difficult or uncertain areas?**  
Which are areas of particular difficulty and/or uncertainty in plan of solving process? Could include deciding:
  - a. What are acceptable levels of uncertainty with which to proceed at various stages?
13. **What info needed?**  
What information is needed to solve the problem? Could include:
  - a. What will be sufficient to test and distinguish between potential solutions?
14. **Priorities?**  
What to prioritize among many competing considerations? What to do first and how to obtain necessary resources?  
Considerations could include: What's most important? Most difficult? Addressing uncertainties? Easiest? Constraints (time, materials, etc.)? Cost? Optimization and trade-offs? Availability of resources? (facilities/materials, funding sources, personnel)
15. **Specific plan for getting information?**  
What is the specific plan for getting additional information? Includes:
  - a. What are the general requirements of a problem-solving approach, and what general approach will they pursue? (These decisions are often made early in the problem-solving process as part of framing.)
  - b. How to obtain needed information? Then carry out those plans. (This could involve many discipline- and problem-specific investigation possibilities such as: designing and conducting experiments, making observations, talking to experts, consulting the literature, doing calculations, building models, or using simulations.)
  - c. What are achievable milestones, and what are metrics for evaluating progress?
  - d. What are possible alternative outcomes and paths that may arise during the problem-solving process, both consistent with predictive framework and not, and what would be paths to follow for the different outcomes?

### Category D. Interpret Information and Choose Solution(s)

This category includes deciding how to analyze, organize, and draw conclusions from available information, reacting to unexpected information, and deciding upon a solution. A biologist

studying aging in worms described how she analyzed results from her experiments, which included representing her results in survival curves and conducting statistical analyses (16, which calculations and data analysis?; 17, how to represent and organize info?), as well as setting up blind experiments (15, specific plan for getting info?) so that she could make unbiased interpretations (18, how believable is info?) of whether a worm was alive or dead. She also described comparing results with predictions to justify the conclusion that worm aging was related to fertility (19, how does info compare to predictions?; 21, appropriate conclusions?; 22, what is best solution?). Deciding how results compared with expectations based on a predictive framework was a key decision that often preceded several other decisions.

Decisions in this category are:

16. **Which calculations and data analysis?**  
What calculations and data analysis are needed? Once determined, these must then be carried out.
17. **How to represent and organize information?**  
What is the best way to represent and organize available information to provide clarity and insights? (Usually this will involve specialized and technical representations related to key features of predictive framework.)
18. **How believable is the information?**  
Is information valid, reliable, and believable (includes recognizing potential biases)?
19. **How does information compare to predictions?**  
As new information comes in, particularly from experiments or calculations, how does it compare with expected results (based on the predictive framework)?
20. **Any significant anomalies?**  
If a result is different than expected, how should one follow up? (This entails first noticing the potential anomaly.) Could involve deciding:
  - a. Does potential anomaly fit within acceptable range of predictive framework(s) (given limitations of predictive framework and underlying assumptions and approximations)?
  - b. Is potential anomaly an unusual statistical variation or relevant data? Is it within acceptable levels of uncertainty?
21. **Appropriate conclusions?**  
What are appropriate conclusions based on the data? (This involves making conclusions and deciding if they are justified.)
22. **What is the best solution?**  
Deciding on best solution(s) involves evaluating and refining candidate solutions throughout the problem-solving process, although they are not always narrowed down to a single solution. May include deciding:
  - a. Which of multiple candidate solutions are consistent with all available information and which can be rejected? (This could be based on comparing data with predicted results.)
  - b. What refinements need to be made to candidate solutions?

### Category E. Reflect

Reflection decisions occur throughout the process and include deciding whether assumptions are justified, whether additional knowledge or information is needed, how well the solution approach is working, and whether potential and then final solutions are adequate. These decisions match the categories of reflection identified by Salehi (2018). A mechanical engineer described developing a model (to inform surgical decisions) of which muscles allow the thumb to function in the most useful manner (22, what is best solution?), including reflecting on how well engineering approximations applied in the biological context (23, assumptions and simplifications appropriate?). He also described reflecting on his approach, that is, why he chose to use cadaveric models instead of mathematical models (25, how well is solving approach working?), and the limitations of his findings in that the “best” muscle identified was difficult to access surgically (26, how good is solution?; 27, broader implications?). Reflection decisions are made throughout the problem-solving process, often lead to reconsidering other decisions, and are critical for success.

Decisions in this category are:

23. **Assumptions and simplifications appropriate?**  
Are previous decisions about simplifications and predictive frameworks still appropriate?
  - a. Do the assumptions and simplifications made previously still look appropriate considering new information?
  - b. Does predictive framework need to be modified?
24. **Additional knowledge needed?**  
Is additional knowledge/information needed? (This is based on ongoing review of one's state of knowledge.) Could involve:
  - a. Is solver's relevant knowledge sufficient?
  - b. Is more information needed and, if so, what?
  - c. Does some information need to be checked? (Is there a need to repeat experiment or check a different source?)
25. **How well is the problem-solving approach working?**  
How well is the problem-solving approach working, and does it need to be modified? This includes possibly modifying the goals. (One needs to reflect on one's strategy by evaluating progress toward the solution.) and reflecting on one's strategy by evaluating progress toward the solution.
26. **How good is the solution?**  
How adequate is the chosen solution? Includes ongoing reflection on potential solutions, as well as final reflection after selecting preferred solution. Can include:
  - a. Decide by exploring possible failure modes and limitations—“try to break” solution.
  - b. Does it “make sense” and pass discipline-specific tests for solutions of this type of problem?
  - c. Does it completely meet the goals/criteria?

### Category F. Implications and Communication of Results

These are decisions about the broader implications of the work, and how to communicate results most effectively. For example,

a theoretical physicist developing a method to calculate the magnetic moment of the muon decided on who would be interested in his work (28, audience for communication?) and what would be the best way to present it (29, best way to present work?). He also discussed the implications of preliminary work on a simplified aspect of the problem (10, approximations and simplifications to make?) in terms of evaluating its impact on the scientific community and deciding on next steps (27, broader implications?; 29, best way to present work?). Many interviewees described that making decisions in this category affected their decisions in other categories.

Decisions in this category are:

**27. Broader implications?**

What are the broader implications of the results, including over what range of contexts does the solution apply? What outstanding problems in the field might it solve? What novel predictions can it enable? How and why might this be seen as interesting to a broader community?

**28. Audience for communication?**

What is the audience for communication of work, and what are their important characteristics?

**29. Best way to present work?**

What is the best way to present the work to have it understood, and its correctness and importance appreciated? How to make a compelling story of the work?

**Category G. Ongoing Skill and Knowledge Development**

Although we focused on decisions in the problem-solving process, the experts volunteered general skills and knowledge they saw as important elements of problem-solving expertise in their fields. These included teamwork and interpersonal skills (strongly emphasized), acquiring experience and intuition, and keeping abreast of new developments in their fields.

Non-decision themes in this category are:

**30. Stay up to date in field**

Staying up to date could include:

- a. Reviewing literature, which does involve making decisions as to which is important.
- b. Learning relevant new knowledge (ideas and technology from literature, conferences, colleagues, etc.)

**31. Intuition and experience**

Acquiring experience and associated intuition to improve problem solving.

**32. Interpersonal, teamwork**

Includes navigating collaborations, team management, patient interactions, communication skills, etc., particularly as how these apply in the context of the various types of problem-solving processes.

**33. Efficiency**

Time management including learning to complete certain common tasks efficiently and accurately.

**34. Attitude**

Motivation and attitude toward the task. Factors such as interest, perseverance, dealing with stress, and confidence in decisions.

**Predictive Framework**

How the decisions were made was highly dependent on the discipline and problem. However, there was one element that was fundamental and common across all interviews: the early adoption of a “predictive framework” that the experts used throughout the problem-solving process. We define this framework as “a mental model of key features of the problem and the relationships between the features.” All the predictive frameworks involved some degree of simplification and approximation and an underlying level of mechanism that established the relationships between key features. The frameworks provided a structure of knowledge and facilitated the application of that knowledge to the problem at hand, allowing experts to repeatedly run “mental simulations” to make predictions for dependencies and observables and to interpret new information.

As an example, an ecologist described her predictive framework for migration, which incorporated important features such as environmental conditions and genetic differences between species and the mechanisms by which these interacted to impact the migration patterns for a species. She used this framework to guide her meta-analysis of changes in migration patterns, affecting everything from her choice of data sets to include to her interpretation of why migration patterns changed for different species. In many interviews, the frameworks used evolved as additional information was obtained, with additional features being added or underlying assumptions modified. For some problems, the relevant framework was well established and used with confidence, while for other problems, there was considerable uncertainty as to a suitable framework, so developing and testing the framework was a substantial part of the solution process.

A predictive framework contains the expert knowledge organization that has been observed in previous studies of expertise (Egan and Greeno, 1974) but goes further, as here it serves as an explicit tool that guides most decisions and actions during the solving of complex problems. Mental models and mental simulations that are described in the naturalistic decision-making literature are similar, in that they are used to understand the problem and guide decisions (Klein, 2008; Mosier *et al.*, 2018), but they do not necessarily contain the same level of mechanistic understanding of relationships that underlies the predictive frameworks used in science and engineering problem solving. While the use of predictive frameworks was universal, the individual frameworks themselves explicitly reflected the relevant specialized knowledge, structure, and standards of the discipline, and arguably largely define a discipline (Wieman, 2019).

**Discipline-Specific Variation**

While the set of decisions to be made was highly consistent across disciplines, there were extensive differences within and across disciplines and work contexts, which reflected the differences in perspectives and experiences. These differences were usually evident in how experts made each of the specific decisions, but not in the choice of which decisions needed to be made. In other words, the solution methods, which included following standard accepted procedures in each field, were very different. For example, planning in some experimental sciences may involve formulating a multiyear construction and data-collection effort, while in medicine it may be deciding on a simple



blood test. Some decisions, notably in categories A, D, and F, were less likely to be mentioned in particular disciplines, because of the nature of the problems. Specifically, decisions 1 (what is important in field?), 2 (opportunity fits solver's expertise?), 27 (broader implications?), 28 (audience for communication?), and 29 (best way to present work?) were dependent on the scope of the problem being described and the expert's specific role in it. These were mentioned less frequently in interviews where the problem was assigned to the expert (most often engineering or industry) or where the importance or audience was implicit (most often in medicine). Decisions 16 (which calculations and data analysis?) and 17 (how to represent and organize info?) were particularly unlikely to be mentioned in medicine, because test results are typically provided to doctors not in the form of raw data, but rather already analyzed by a lab or other medical technology professional, so the doctors we interviewed did not need to make decisions themselves about how to analyze or represent the data. Qualitatively, we also noticed some differences between disciplines in the patterns of connections between decisions. When the problem involved development of a tool or product, most commonly the case in engineering, the interview indicated relatively rapid cycles between goals (3), framing problem/potential solutions (8), and reflection on the potential solution (26), before going through the other decisions. Biology, the experimental science most represented in our interviews, had strong links between planning (15), deciding on appropriate conclusions (21), and reflection on the solution (26). This is likely because the respective problems involved complex systems with many unknowns, so careful planning was unusually important for achieving definitive conclusions. See Supplemental Text and Supplemental Table S2 for additional notes on decisions that were mentioned at lower frequency and decisions that were likely to be interconnected, regardless of field.

## DISCUSSION

This work has created a framework of decisions to characterize problem solving in science and engineering. This framework is empirically based and captures the successful problem-solving process of all experts interviewed. We see that several dozen experts across many different fields all make a common set of decisions when solving authentic problems. There are flexible linkages between decisions that are guided by reflection in a continually evolving process. We have also identified the nature of the “predictive frameworks” that S&E experts consistently use in problem solving. These predictive frameworks reveal how these experts organize their disciplinary knowledge to facilitate making decisions. Many of the decisions we identified are reflected in previous work on expertise and scientific problem solving. This is particularly true for those listed in the planning and interpreting information categories (Egan and Greeno, 1974). The priority experts give to framing and planning decisions over execution compared with novices has been noted repeatedly (e.g., Chi *et al.*, 1988). Expert reflection has been discussed, but less extensively (Chase and Simon, 1973), and elements of the selection and implications and communication categories have been included in policy and standards reports (e.g., AAAS, 2011). Thus, our framework of decisions is consistent with previous work on scientific practices and expertise, but it is more

complete, specific, empirically based, and generalizable across S&E disciplines.

A limitation of this study is the small number of experts we have in total, from each discipline, and from underrepresented groups (especially lack of female representation in engineering). The lack of randomized selection of participants may also bias the sample toward experts who experienced similar academic training (STEM disciplines at U.S. universities). This means we cannot prove that there are not some experts who follow other paths in problem solving. As with any scientific model, the framework described here should be subjected to further tests and modifications as necessary. However, to our knowledge, this is a far larger sample than used in any previous study of expert problem solving. Although we see a large amount of variation both within and across disciplines in the problem-solving process, this is reflected in how experts make decisions, not in what decisions they make. The very high degree of consistency in the decisions made across the entire sample strongly suggests that we are capturing elements that are common to all experts across science and engineering. A second limitation is that decisions often overlap and co-occur in an interview, so the division between decision items is often somewhat ambiguous and could be defined somewhat differently. As noted, a number of these decisions can be interconnected, and in some fields are nearly always interconnected.

The set of decisions we have observed provides a general framework for characterizing, analyzing, and teaching S&E problem solving. These decisions likely define much of the set of cognitive skills a student needs to practice and master to perform as a skilled practitioner in S&E. This framework of decisions provides a detailed and structured way to approach the teaching and measurement of problem solving at the undergraduate, graduate, and professional training levels. For teaching, we propose using the process of “deliberate practice” (Ericsson, 2018) to help students learn problem solving. Deliberate practice of problem solving would involve effective scaffolding and concentrated practice, with feedback, at making the specific decisions identified here in relevant contexts. In a course, this would likely involve only an appropriately selected set of the decisions, but a good research mentor would ensure that trainees have opportunities to practice and receive feedback on their performance on each of these 29 decisions. Future work is needed to determine whether there are additional decisions that were not identified in experts but are productive components of student problem solving and should also be practiced. Measurements of individual problem-solving expertise based on our decision list and the associated discipline-specific predictive frameworks will allow a detailed measure of an individual's discipline-specific problem-solving strengths and weaknesses relative to an established expert. This can be used to provide targeted feedback to the learner, and when aggregated across students in a program, feedback on the educational quality of the program. We are currently working on the implementation of these ideas in a variety of instructional settings and will report on that work in future publications.

As discussed in the *Introduction*, typical science and engineering problems fail to engage students in the complete problem-solving process. By considering which of the 29 decisions are required to answer the problem, we can more clearly

articulate why. The biology problem, for example, requires students to decide on a predictive framework and access the necessary content knowledge, and they need to decide which information they need to answer the problem. However, other decisions are not required or are already made for them, such as deciding on important features and identifying anomalies. We propose that different problems, designed specifically to require students to make sets of the problem-solving decisions from our framework, will provide more effective tools for measuring, practicing, and ultimately mastering the full S&E problem-solving process.

Our preliminary work with the use of such decision-based problems for assessing problem-solving expertise is showing great promise. For several different disciplines, we have given test subjects a relevant context, requiring content knowledge covered in courses they have taken, and asked them to make decisions from the list presented here. Skilled practitioners in the relevant discipline respond in very consistent ways, while students respond very differently and show large differences that typically correlate with their different educational experiences. What apparently matters is not what content they have seen, but rather what decisions they have had practice making. Our approach was to identify the decisions made by experts, this being the task that educators want students to master. Our data do not exclude the possibility that students engage in and/or should learn other decisions as a productive part of the problem-solving process while they are learning. Future work would seek to identify decisions made at intermediate levels during the development of expertise, to identify potential learning progressions that could be used to teach problem solving more efficiently. What we have seen is consistent with previous work identifying expert–novice differences but provides a much more extensive and detailed picture of a student's strengths and weaknesses and the impacts of particular educational experiences. We have also carried out preliminary development of courses that explicitly involve students making and justifying many of these decisions in relevant contexts, followed by feedback on their decisions. Preliminary results from these courses are also encouraging. Future work will involve the more extensive development and application of decision-based measurement and teaching of problem solving.

## ACKNOWLEDGMENTS

We acknowledge the many experts who agreed to be interviewed for this work, M. Flynn for contributions on expertise in mechanical engineering, and Shima Salehi for useful discussions. This work was funded by the Howard Hughes Medical Institute through an HHMI Professor grant to C.E.W.

## REFERENCES

- ABET. (2020). *Criteria for accrediting engineering programs, 2020–2021*. Retrieved November 23, 2020, from [www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2020-2021](http://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2020-2021)
- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014). Control of gene expression. In *Molecular Biology of the Cell* (6th ed., pp. 436–437). New York: Garland Science. Retrieved November 12, 2020, from <https://books.google.com/books?id=2xIwDwAAQBAJ>
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC: Retrieved February 12, 2021, from <https://visionandchange.org/finalreport>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, 6(3), 271–315. [https://doi.org/10.1207/s15327809jls0603\\_1](https://doi.org/10.1207/s15327809jls0603_1)
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. [https://doi.org/10.1207/s15516709cog0502\\_2](https://doi.org/10.1207/s15516709cog0502_2)
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Clemmons, A. W., Timbrook, J., Herron, J. C., & Crowe, A. J. (2020). BioSkills Guide: Development and national validation of a tool for interpreting the *VisionandChange* core competencies. *CBE—Life Sciences Education*, 19(4), ar53. <https://doi.org/10.1187/cbe.19-11-0259>
- Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT Press.
- Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, 21(1), 49–58. [https://doi.org/10.1016/S0193-3973\(99\)00050-7](https://doi.org/10.1016/S0193-3973(99)00050-7)
- Egan, D. E., & Greeno, J. G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving in L. In Gregg, W. (Ed.), *Knowledge and cognition*. Potomac, MD: Erlbaum.
- Ericsson, K. A. (2018). The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts. In Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 745–769). Cambridge, United Kingdom: Cambridge University Press. <https://doi.org/10.1017/9781316480748.038>
- Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R., (Eds.) (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge, United Kingdom: Cambridge University Press.
- Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. A., (Eds.) (2018). *The Cambridge handbook of expertise and expert performance* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. <https://doi.org/10.1037/0033-295X.87.3.215>
- Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, 101(2), 336–342. doi: 10.1037/0033-295X.101.2.336
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In Stevenson, H. W., Azuma, H., & Hakuta, K. (Eds.), *A series of books in psychology. Child development and education in Japan* (pp. 262–272). New York: Freeman/Times Books/Henry Holt.
- Heckler, A. F. (2010). Some consequences of prompting novice physics students to construct force diagrams. *International Journal of Science Education*, 32(14), 1829–1851. <https://doi.org/10.1080/09500690903199556>
- Heller, J. I., & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. *Cognition and Instruction*, 1(2), 177–216. [https://doi.org/10.1207/s1532690xci0102\\_2](https://doi.org/10.1207/s1532690xci0102_2)
- Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, 60, 627–636. <https://doi.org/10.1119/1.17117>
- Hsieh, H-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Huffman, D. (1997). Effect of explicit problem-solving instruction on high school students' problem-solving performance and conceptual understanding of physics. *Journal of Research in Science Teaching*, 34(6), 551–570. [https://doi.org/10.1002/\(SICI\)1098-2736\(199708\)34:6<551::AID-TEA2>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1098-2736(199708)34:6<551::AID-TEA2>3.0.CO;2-M)
- Kind, P., & Osborne, J. (2016). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 10(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50(3), 456–460.
- Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences*, 9(2), 105–143.
- Kuo, E., Hallinen, N. R., & Conlin, L. D. (2017). When procedures discourage insight: Epistemological consequences of prompting novice physics

- students to construct force diagrams. *International Journal of Science Education*, 39(7), 814–839. <https://doi.org/10.1080/09500693.2017.1308037>
- Larkin, J., & Reif, F. (1979). Understanding and teaching problem-solving in physics. *European Journal of Science Education*, 1(2), 191–203. doi: 10.1080/0140528790010208
- Lintern, G., Moon, B., Klein, G., & Hoffman, R. (2018). Eliciting and representing the knowledge of experts. In Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed). (pp. 165–191). Cambridge, United Kingdom: Cambridge University Press.
- Mosier, K., Fischer, U., Hoffman, R. R., & Klein, G. (2018). Expert professional judgments and “naturalistic decision making.” In Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed). (pp. 453–475). Cambridge, United Kingdom: Cambridge University Press.
- National Research Council (NRC). (2012a). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- NRC. (2012b). Problem solving, spatial thinking, and the use of representations in science and engineering. In *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (pp. 75–118). Washington, DC: National Academies Press. <https://doi.org/10.17226/13362>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>
- Organisation for Economic Cooperation and Development. (2019). PISA 2018 science framework. In *PISA 2018 assessment and analytical framework* (pp. 97–117). Paris: OECD Publishing. <https://doi.org/10.1787/f30da688-en>
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
- Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., & Upmeier Zu Belzen, A. (2020). A framework to foster problem-solving in STEM and computing education. *Research in Science & Technological Education*, 38(1), 105–130. <https://doi.org/10.1080/02635143.2019.1600490>
- Quacquarelli Symonds. (2018). *The global skills gap in the 21st century*. Retrieved July 20, 2021, from [www.qs.com/portfolio-items/the-global-skills-gap-in-the-21st-century/](http://www.qs.com/portfolio-items/the-global-skills-gap-in-the-21st-century/)
- Salehi, S. (2018). *Improving problem-solving through reflection (Doctoral dissertation)*. Stanford Digital Repository, Stanford University. Retrieved February 18, 2021, from <https://purl.stanford.edu/gc847wj5876>
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.
- Simon, H. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4(3–4), 181–201. [https://doi.org/10.1016/0004-3702\(73\)90011-8](https://doi.org/10.1016/0004-3702(73)90011-8)
- Singh, C. (2002). When physical intuition fails. *American Journal of Physics*, 70, 1103–1109. <https://doi.org/10.1119/1.1512659>
- Wayne State University. (n.d). *Mechanical engineering practice qualifying exams. Wayne State University Mechanical Engineering department*. Retrieved February 23, 2021, from [https://engineering.wayne.edu/me/exams/mechanics\\_of\\_materials\\_-\\_sample\\_pqe\\_problems\\_.pdf](https://engineering.wayne.edu/me/exams/mechanics_of_materials_-_sample_pqe_problems_.pdf)
- Wieman, C. E. (2015). Comparative cognitive task analyses of experimental science and instructional laboratory courses. *Physics Teacher*, 53, 349–351. <https://doi.org/10.1119/1.4928349>
- Wieman, C. E. (2019). Expertise in university teaching & the implications for teaching effectiveness, evaluation & training. *Daedalus*, 148(4), 47–78. [https://doi.org/10.1162/daed\\_a\\_01760](https://doi.org/10.1162/daed_a_01760)
- Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22(3), 319–346. [https://doi.org/10.1016/S0364-0213\(99\)80043-3](https://doi.org/10.1016/S0364-0213(99)80043-3)