Sophisticated abstract thought is perhaps humanity's greatest asset, and language its most important technology[1]. At their best, they serve individual and collective survival and wellbeing, guide our explorations of reality, and immeasurably amplify our potential as a species. But abstract thought begs an ever-present question: to what extent is our thought—its contents, its characteristic patterns—aligned with that which we call reality? To what extent is it a trustworthy guide in our explorations of reality, reliably pointing us towards genuine problems that will richly reward our attention, and guarding us from pseudo-problems? How can we determine where and how thought is misaligned, and how to best correct it? This is the alignment problem, and perhaps one of humanity's greatest enigmas.

Today, we look upon modern science as an exemplar of a social activity that creates abstract systems of thought, expressed in natural and formal languages, which mirror physical reality with uncanny precision, which articulate profound regularities in natural processes and accurately describe the nature of the microscopic constituents of matter.

Due to its extraordinary track record of discoveries since the 1650s, and the transformative technologies based upon them, modern science has a central place in the culture of today's society. In particular, the mechanistic conception of reality on which it was founded is accorded a particular authority. Today, most of us are taught that physical reality really does consist of inanimate matter composed of atoms and subatomic particles of various types; is really governed by universal laws; that living things are essentially complex arrangements of inanimate matter; and so forth. And this conception is reflected back to us daily by the culture that surrounds us. Although developments in theoretical physics—particularly quantum theory—have challenged most of the key assumptions that underpin the mechanical conception, this conception continues to hold sway, serving as a lens through which we look at the physical world.

However, I believe that the great successes of modern science have blunted our sensitivity to the alignment problem. I believe that we—as scientists—have become complacent: we tacitly believe that the mechanistic conception is more or less right, modulo corrections that only matter in the microscopic realm, which are in any case are taken care of as long as we make judicious use of the quantum formalism when it's appropriate to do so. But, there is another possibility which I contend it behoves us to seriously consider: that the mechanistic conception is merely one lens—albeit an extraordinarily fruitful one—through which to look at reality; that there are other lenses which will reveal other facets of reality that the mechanical conception presently obscures from view; and that not being alive to these other possibilities is now inhibiting the evolution of human thought and, by extension, holding back the development of science.

It is my contention that, to safeguard the creative flexibility of human thought and modern science, it is essential at this moment in history to methodically and critically reflect upon and clarify the metaphysical underpinnings of modern science, and sympathetically consider metaphysical views—both western and non-western—in light of the products of modern science and its open questions. In particular, it is essential that we complete the task of developing of a coherent con-

[1] On language as technology, see e.g. [1].

ception of reality that does justice to the full content of quantum theory, as this provides the clearest opportunity for articulating a precise alternative to the mechanical conception which is deeply informed by a battle-tested physical theory, and for becoming fully aware of the the deeply-engrained assumptions implicit in the mechanical conception.

However, I believe that modern science—as it is currently institutionally organized, practiced, and taught—is too far skewed to the utilitarian and pragmatic to effectively engage in such a task. In particular, I believe that science as we know it must undergo a thorough-going transformation which reconnects it with the reflective and contemplative modes of being, both at the level of the individual scientist and at the level of the intellectual disciplines that exemplify these modes of being. Below, I discuss the prospects of such transformation and how scientists so inclined can take part in such a transformation; and why this is of importance not only for the future of science, but for the health and well-being society as a whole.

I. EMERGENCE OF THE ALIGNMENT PROBLEM, AND ITS AMELIORATION

In the nomadic mode of living, language enabled coordination of the tribe in daily affairs, and mostly referred to the tangible objects and events in immediate experience, as well as to recent memories of everyday events and to plans to be enacted in the immediate future. Language may have also referred to hypothetical natural forces or deities, but the details would have been relatively unimportant—what would have mattered much more is whether these notions supported the cohesion and harmony of the tribe.

However, the transition to the agricultural mode of living placed new demands on thought and language [2]. The need for coordination between a vastly greater number of people—most of whom were not personally known to one another—and the greatly expanded temporal horizon of memory and planning required by agriculture, forced thought and language to become ever more abstract, detailed, and precise. In particular, thought began to embody an abstract model of physical reality that extrapolated far beyond the immediate objects and events of perception, and whose faithfulness to reality became increasingly important to the survival and thriving of the agricultural communities. This, in turn, increasingly brought the alignment problem to the fore.

Humanity has devised many methods to ameliorate the alignment problem. Accurate record-keeping by trusted individuals safeguards collective memory. Systems of laws emphasize conflict-resolution in an evidence-based manner. Social norms of truthful speech disfavour factual distortion. Special disciplines—such as philosophy and metaphysics—train individuals to slow down and step back from the stream of thinking to become aware of implicit assumptions and errors; to recognize the gap between thought, word, and lived experience; and to be more alive to alternative patterns of thought. Spiritual practices train individuals to observe moods, emotions, and intentions; to notice how streams of thought are initiated and how they are directed by these mental forces; and to purify the heart-mind of habits and intentions that distort the thinking process.

II. THE DISTINCTIVE CHARACTER OF MODERN SCIENCE

Scientific activity is a particular means of addressing the alignment problem. In its broadest sense, science is a social activity devoted to the development of abstract models of chosen aspects of physical reality for the betterment of the collective. As such, it seeks to harness the exploratory mode of being, with its delight in uncovering the hitherto hidden and mysterious, in the service of the utilitarian mode, with its drive to act and plan so as to maintain the worldly survival, thriving, and wellbeing of the collective.

Science has arisen in sophisticated forms in numerous cultures—in Ancient Greece, China, the Arabic world. If anything distinguishes modern science ($\sim$1600–present) from these precursors, it is the breadth and audacity of its vision: that beneath the stream of ever-changing, diverse sensory experience lies a unitary mathematical order; and that there is a reliable method by which we can gain access to that order. In particular, that by constructing mathematical theories of particular types of natural phenomena, we can quantitatively connect some of our forebears' metaphysical speculations concerning the nature of physical reality to that which is accessible to the senses. Through quantitative testing of these theories in controlled experiments, we can weed out incorrect hypotheses, and hone in on ever more general mathematical laws which embody and precisely articulate general principles that regulate natural processes.

The founding myth of modern science appeals simultaneously to many different modes of being. In its promise to connect lived experience rigorously to the realm of metaphysical thought, it not only synthesizes the exploratory and utilitarian modes of being, but also the reflective mode characteristic of philosophical thinking. In its posit of the existence of a unitary mathematical order and general principles of nature, modern science appeals to the contemplative mode, with its desire to reflect in thought the deepest fabric of reality. In its posit of a scientific method by which this order can come to be known, it appeals to the utilitarian mode: we mortals—as finite and limited as we may be—can legitimately aspire to potentially limitless knowledge of the mathematical order that underpins the physical universe, and enjoy the benefits of the unprecedented level of control over natural processes that this knowledge entails.

Over the course of the seventeenth century, modern science articulated a specific metaphysical conception of nature—an atomistic mechano-geometric conception, or mechanical conception for short—that weaved together Democritus' atomism and Euclidean geometry, bequeathed by Ancient Greece, with a novel notion of mechanical causation (replacing Aristotle's theory of causation)[2]. In essence: all matter is composed of elementary parts (variously atoms or elements) that compose the complex objects—both living and nonliving, earthly and heavenly—in the world around us. These atoms are *individuals* (in the metaphysical sense of possessing independent existence and formal unity), exist in space, possess properties that are exhaustively quantifiable, interact via local influences, and move in strict accordance with universal mathematical laws.

This metaphysical conception was in harmony with the notion of controlled experiment articu-

[2] On the mechanical conception of classical physics, and its relation to the scientific method, see e.g. [3].

lated by Bacon: as matter consists of individual parts, each subject to the laws of nature, one can reliably build up a knowledge of the whole by studying the parts; as matter is subject to mechanical causation and its properties are quantifiable, one can isolate such parts from their past history or future happenings; since interactions of parts are local, one can insulate these parts to a considerable degree from contemporaneous events occurring elsewhere. Alternative metaphysical views were aired and discussed—nonmechanical theories of causation; propensities as distinct from properties; monistic (as opposed to pluralistic) views of matter; and vital forces emanating from living things (as distinct from universal mathematical laws). But the mechanical conception attracted ever-increasing support due to the successes of the physical discoveries—due to Galileo, Huygens, Newton, and others—based upon it.

III. SHIFT TOWARDS UTILITARIANISM

The seventeenth century saw ample evidence of the fecundity of modern science—of its mechanistic conception, and the scientific method. But the great social utility of the creative products of modern science became apparent over the course of the eighteenth and particularly nineteenth centuries, which supported the first and second Industrial Revolutions.

The social utility of science transformed the nature of scientific activity. This transformation took place in two distinct shifts, the first gradually, the second suddenly. The first occurred roughly during the period 1700–1930, over which period scientific discoveries and technological inventions transformed everyday life and restructured the global order through conquest and wars fuelled by these inventions. The second shift took place suddenly during $\sim$1930–1950, as a result of the massive social upheavals in Europe due to the Great Depression and World War II.

Each of these shifts affected both the organization of the scientific activity, and the style of work carried out. The first shift led to scientific activity being increasingly brought into the academy, supported by learned societies and periodicals, with increasing governmental support. The second shift led to a displacement of the centre of gravity of scientific research from Europe to the Anglo-Saxon World (particularly the US), and an intensification of governmental and industrial support for scientific activity.

Both shifts led to a greater emphasis on scientific activity that reliably delivers tangible results and predictive control of natural processes for the benefit of humankind, as opposed to exploratory and particularly reflective activity farther removed from such desired outcomes. For example, during the first shift, there was a hardening of the metaphysical presumptions—an increasing intolerance for alternative metaphysical positions within science, and a marginalization of philosophical reflection, reinforced by bold public-facing pronouncements advertising a mechanical world view, backed by the accomplishments of modern science, as a replacement for conceptions of reality favoured by other compartments of culture. In the second shift, the relatively utilitarian Anglo-Saxon culture placed far greater emphasis on tangible, useful outcomes, and actively discouraged philosophical reflection within science; and the emergence of a competitive grant and rapid-publication culture further disfavoured exploratory activity.

IV. WHY IS THE UTILITARIANISM PROBLEMATIC?

A utilitarian-oriented science excels in the theoretical development and technological applications of the known. Quantum field theory and the standard model of particle physics continue the mathematical patterns established by quantum theory in the 1920s. The astounding technological achievements of the last hundred years have transformed everyday life out of all recognition, yet are largely founded on theories—electromagnetism and quantum theory—that were in place by the mid-1860s and mid-1920s, respectively.

But a utilitarian-oriented science is also averse to going where the light does not shine brightly—into what has not yet been formalized, into the implicit, to what shades over to the philosophical or—god forbid—the metaphysical or spiritual. As a result, it does not cultivate new soil so that it may serve as fertile ground in which the seeds of genuinely new scientific growth may germinate; and lets new seeds slip through its hands, even as they appear as the result of it own activity. As a result, it risks stagnation, even in the midst of the apparent splendor of its technological fruits.

There is perhaps no clearer example of this aversion than the response of the physics community to the challenges posed by quantum theory. As is well-known, quantum theory profoundly challenges the gamut of concepts that constitute the mechanical conception—the notion of particle as individual existent is challenged by the behaviour of systems of identical quantum particles; the spatial-groundedness of all properties by nonlocal properties in entangled states; mechanistic causation by probabilistic quantum predictions; passive measurability by the quantum measurement postulate; and so forth.

Many of the founders of quantum theory—such as Bohr, Heisenberg, Schroedinger, and Pauli—devoted considerable efforts to reflect tease out the hidden implications of quantum theory (see e.g. [4–9]). Their efforts gave rise to such notions as complementarity, entanglement, steering, and to the EPR paradox. Most also reflected philosophically on the implications of quantum theory for the nature and limitations of the patterns of thought characteristic of the mechanistic conception, and on the relationship between physics and other compartments of culture (such as psychology and philosophy). However, such foundational probing and philosophical reflection amongst physicists faded out following the second utilitarian shift—in the Anglo-Saxon world, such activity was generally regarded as tangential to the scientific enterprise.

However, such foundational reflections did not cease entirely. And there is perhaps no clearer illustration of the scientific value of such reflection than the emergence of the fields of quantum information and quantum computation, which sprang from the foundational work of physicists such as Bohm (in his reformulation of quantum theory [10]) and Bell [11], who managed to sustain the spirit of foundational enquiry. These fields have now not only extensively explored facets of the quantum world that, in the 1930s, seemed mysterious and (in the case of entanglement) perhaps mere artefacts of the newly-created quantum formalism, but have shown how they can be technologically harnessed in ways that will bring forth a second wave of quantum technologies in the coming decades.

But that still leaves the question whether science—for example theoretical physics—as a whole can prosper without seriously engaging in foundational and philosophical reflection. I believe that while it may prosper for some time, the lack of such foundations will ultimately prove a limiting factor. For example, examination of the historical developments that preceeded major steps in theoretical physics shows that foundational thinking was often pivotal. For example, Mach's foundational critique of Newtonian mechanics in 1883 [12]—of a theory that had then been in existence for almost 200 years and was widely accepted as the basis of modern physics—profoundly influenced Einstein's path to his theories of relativity.

Moreover, the last two decades are remarkable for the publication of a series of detailed critiques of major research programs in theoretical physics—string theory [13], theoretical and particle physics [14–16], and standard cosmology [17]—which suggest the entrenchment of suboptimal research programs and the neglect of promising rival research programs. But other adverse consequences cannot be so easily pointed to—new paths of exploration and reflection that were never taken by researchers in favour of well-trodden paths that will yield more predictable, publishable outcomes; young creative minds that turned away from physics at an early stage because their curious questions do not seem welcome in a teaching context which emphasizes problem solving over foundational understanding and reflection.

V. TRANSFORMING SCIENCE

A healthy science is one that integrates the reflective and contemplative modes of being. Historically, the reflective mode has been particularly cultivated in philosophy. Philosophy has furnished most of the ideas that were formalized in the early theories of modern science. Democritus' atomism was once hypothetical metaphysical fancy, yet some 2500 years later is broadly regarded (whether rightly or wrongly is another matter) as experimentally and theoretically established fact. Mach's critique of the Newtonian conceptions of absolute space and absolute rotation, and more broadly his operational view of physical theories, can be traced back at least partially to Hume's penetrating philosophical analysis [18] of the fundamental concepts, such as cause and existence of external persistent objects, which are implicit both in common speech about the everyday physical world and in classical physics.

The contemplative mode has been historically cultivated in the world's spiritual traditions, but at least some of the fruits of these traditions is now widely accessible in the form of secularized techniques[3], many of which have been studied in an academic setting. From the viewpoint of the development of human thought, contemplative technique train the individual to step outside the stream of thinking and treat thought as an object to be observed without getting caught up in its momentum; and also to suspend thinking activity entirely whilst remaining fully aware in all other respects (of intention, mood, emotion, sensation). Such techniques thereby cultivate the ability to recognize the gaps between the words used to label and describe experience, and the

[3] One of the most well-known techniques is Mindfulness-based Cognitive Therapy (MBCT) (see e.g. https://www.mbct.com).

experience itself; to recognize what implicit assumptions are brought into play when we label and describe experience; and to see how the act of thinking about an experience changes or colours that experience.

Together, the reflective and contemplative modes act as a powerful counterbalance to the natural human tendency to reify well-established patterns of thought that have proven effective in the past.

A. General Conditions for Transformations

Precisely what will or could transform science so as to effect such an integration of the reflective and contemplative modes is a quite different matter. Transformation is obvious in hindsight. But when one is in the midst of transformation, especially in the early stages, its stirrings are very hard to identify. Transformation requires general supportive conditions as well as specific unequivocal demonstrations of the fecundity of a new way of approaching open scientific questions.

A number of general supporting conditions for such transformation do appear to exist. First, the second utilitarian shift took place some 70 years ago, at a point when the Anglo-Saxon world was in the ascendency, Europe was in the process of rebuilding, and Asia was still reeling from the after-effects of WWII. Since that time, European culture has entirely re-established itself, and exhibits growing self-confidence; and scientists across Asia (in countries such as China and India, themselves increasingly self-expressive on the world stage) are now part of a global scientific community. Most of these cultures have distinctive and rich philosophical and contemplative traditions, and many regard science as merely one voice of reason and authority amongst many when it comes to the nature of reality.

Second, over this same timeframe, considerable scholarly work has been carried out which makes available the past achievements and detailed scientific and philosophic thought of other cultures. For example, Joseph Needham's massive "Science and Civilization in China" project has brought together scholars from the West and China to produce a multi-volume compilation that describes the scientific and philosophic contributions of China. Similarly, there is an active tradition of scholarly translation and analysis of metaphysical systems of Indian philosophy (such as Madhyamaka), as well as efforts to surface and analyse the scientific and philosophical of the Arabic world.

Third, in various scientific disciplines, there have been notable movements to break out of the constraints characteristic of mechanistic patterns of thinking. One common thread is the desire to understand the complex emergent behaviour of systems consisting of a large number of interacting agents. This has given rise, for instance, to ecological thinking, and to interdisciplinary research into complex systems (see e.g. [19]). The latter aims to break away from the reductive thinking characteristic of mechanical philosophy, and instead embrace the messy complexity characteristic of living systems—collections of non-ideal economic agents; ant colonies; neural structures subject to Hebbian learning; gene networks—in the hope of uncovering general rules of emergent behaviour.

Fourth, as mentioned earlier, there is increasing disquiet in theoretical physics that conventional research programs are too conservative, and that deeper foundational understanding of our best physical theories is crucial for the development of the subject. For example, over the last few decades, the field of foundations of quantum theory has been greatly stimulated by the prospect of understanding quantum theory by reconstructing (i.e. systematically deriving) the mathematical formalism of the theory from new principles that are information-theoretic in nature. Several detailed reconstructions of quantum theory now exist, and these distill the detailed content of abstract quantum formalism in the form of postulates that are much more amenable to philosophical reflection.

Fifth, compared with the 1960s, the publication and funding landscape has considerably diversified. Open access publications and online archives have greatly expanded the diversity of publication venues, and the funding landscape includes private foundations (such as FQxI, Templeton Foundation, Fetzer Institute, and many others) which seek to foster more visionary, higher-risk research. In addition, from time to time, one witnesses the emergence of institutes, such as the Santa Fe Institute and Perimeter Institute, with visionary charters that offer crucial support for scientists who wish to pursue less well trodden paths.

B. A Specific Open Question: Formulation of a Quantum Conception of Reality

There are potentially many specific actions that could be taken that demonstrate the fecundity of a more reflective, contemplative approach to science. I will elaborate on just one such action.

As mentioned earlier, a major open task in physics since the creation of quantum mechanics is the construction of a conception of reality that does justice to the mathematical formalism of these theories. To my mind, it is essential that such a conception—a quantum conception of reality—be constructed, so that scientists may be in possession of a richly detailed alternative to mechanical conception, based on our most powerful physical theory.

It is important to emphasize that although many so-called interpretations of quantum theory presently exist, they depend on very little of the formalism of the theory itself. In that sense, they are akin to loose-fitting clothes draped over the quantum formalism. Rather, what is sought here is a metaphysically-precise conception that closely fits the contours of the formalism[4], so that we can say explicitly which of the postulates of the mechanical conception can be retained, which must be modified (and why), and which must be abandoned.

Over the last twenty years, the reconstruction program has yielded a number of detailed derivations of many parts of the quantum formalism from relatively simpler physically-motivated postulates. These set the stage for the next step, namely the development of a metaphysically well-informed conception of reality that makes sense of the postulates from select reconstructions. In this task, one will need to bring to bear all the relevant metaphysics.

[4] For more details on the quantum reconstruction program, and the contrast with interpretations, see for example [20–22].

If this interpretative stage were successfully completed, we would be in possession of a precise metaphysical conception stated in natural language which captures, in essence, the content of quantum theory. As such, this conception would be broadly accessible not only to scientists, but to nonscientists in the academy and to those in the broader society who are intrigued by modern science.

Furthermore, if the deepened understanding provided by this conception were to feed back to theoretical physics, so as to suggest new approaches to theory-building (for example, new approaches to quantum gravity), or new ways to harness quantum theory itself (rather as the informational perspective on quantum theory spawned quantum cryptography and quantum computation), it would potentially generate an even stronger interest in this way of doing physics.

For example, one open question is how to theoretically understand the entanglement of identical particles, a question that has become of interest in connection to quantum technologies. The essential theoretical difficulty is that, on the standard reading of the quantum formalism, a symmetrized state of two identical particles satisfies the criterion for entanglement. Yet it seems obvious that two electrons in separate labs, and prepared independently, cannot be entangled. Although various resolutions of this problem have been proposed in the literature, none are entirely satisfactory; and the enduring difficulty may be traced to an inadequate understanding of the nature of identical quantum particles themselves. A deeper understanding of the quantum rules for handling identical particles may illuminate or resolve this open question.

VI. SOCIAL IMPORTANCE OF TRANSFORMING MODERN SCIENCE

Human society is at a uniquely perilous moment. Due to its astonishing sequence of discoveries and the impact of the technologies based upon them, science enjoys unique social support and prestige. As such, scientists are in a unique position to guide humanity through what awaits it in the coming centuries. A science that is more integrated, more balanced, more enlightened, will be a better, wiser, guide.