Contents

Editorial 1
Tai-Danae Bradley

The Queen of the Sciences: Reclaiming the Rightful Place of Theology and Creation 3
Abner Chou

A New Perspective of Entropy 35
Tai-Danae Bradley

Imitation Learning: the Machine Learning Version of Discipleship 67
Monica B. Vroman

Christ in Creation 91
John F. MacArthur
This issue launches The Journal of The Math3ma Institute, a new home for expository articles on research in science, technology, engineering, and mathematics (STEM) written for the ambitious layperson with interests in these fields of study. Prospective contributors to the journal include faculty members of The Master’s University as well as scholars and research scientists from other institutions who wish to make their work accessible to broad audiences.

The current issue, for example, contains articles on original peer-reviewed research in both computer science and pure mathematics targeted at non-experts. *Imitation Learning* by Dr. Monica B. Vroman gives a friendly introduction to and overview of machine learning, a subfield of artificial intelligence, and hones in on a particular framework called imitation learning. The content of the article is based on Dr. Vroman’s PhD thesis and the article itself is an accessible invitation to her scholarly work in this field. My article *A New Perspective of Entropy* similarly seeks to distill peer-reviewed research-level mathematics for a wide audience. It describes a new connection between information theory and two branches of higher mathematics known as abstract algebra and topology and details a new way to understand entropy from this perspective. Both articles are intended to illuminate and demystify interesting, yet technical, scientific work for the curious non-expert.

This naturally leads to the question of why. Why make an effort to distill complex scientific ideas for a broad audience? Our motivation is grounded in a biblical worldview and a Christ-centered perspective, which is a pillar of The Math3ma Institute (TMI), founded in Fall 2021 at The Master’s University. At TMI, we cling to the truth that in Christ all things were created (Colossians 1:17), that He upholds all things by the word of His power (Hebrews 1:3), and that any endeavor to explore the wonders of creation finds its deepest and richest fulfillment in the acknowledgment of God as Creator (Genesis 1:1). At the same time, we understand the...
reality that the beauty and wonder of such discoveries are often inaccessible to lay audiences due to the high level of technicality and specialized language. As a result, those who are initially eager to engage in the sciences oftentimes become discouraged from doing so and even from entering the field altogether. These barriers may unintentionally convey the idea that STEM belongs to a select few and only to them. By communicating scholarly scientific work in engaging and accessible ways, we may help to break down the barriers that typically prevent others from seeing the simplicity and beauty of scientific discovery.

Accordingly, the journal also welcomes articles on Christian theology pertaining to STEM. This inaugural issue opens with The Queen of the Sciences by Dr. Abner Chou, for example, who expounds upon the doctrine of creation and its role in reclaiming theology as the queen of the sciences. The editors may further accept shorter submissions such as relevant book reviews or featured articles from select scholars. For instance, this present issue concludes with Christ in Creation, a featured article by Dr. John MacArthur that connects the Genesis account of creation to the New Testament with a special emphasis on the central role of Christ, the Word, detailed in John 1:1–5.

The journal thus differs from existing academic outlets in its writing style (which is explanatory), its scope (which embraces all of STEM), and its motivation for both. It is our prayer that every publication from The Journal of The Math3ma Institute would amplify the joy and hope available when the sciences are approached with the motivations described above.
The Queen of the Sciences: Reclaiming the Rightful Place of Theology and Creation

Dr. Abner Chou
PhD, Old Testament, The Master’s Seminary
John F. MacArthur Endowed Fellow and Interim President at The Master’s University

Abstract: Historically, theology was viewed as the queen of the sciences. But in recent days this has fallen out of favor, especially due to the unpopularity of the doctrine of creation. Instead, science is viewed as its own autonomous foundation. This article surveys through the issues surrounding creation and argues that a realism of biblical authority and revelation establishes theology and creation as a necessary framework for science. It also will contend that the interpretation of Genesis 1–3 is clear and clearly historical as well as that the doctrine of creation is inextricably linked with the totality of Christian theology. Even more, it will survey God’s plan of redemption and illustrate that creation is the basis and driver of God’s redemptive work. Creation holds the answers to the toughest questions people have about this world and evil. With that, by virtue of divine authority, theology is the queen of the sciences, and within this, the doctrine of creation helps to restore the true value and beauty of science. Therefore, it should be the starting point of the sciences.

It is fitting to begin this inaugural issue with a discussion on creation. As Scripture states, creation is “in the beginning” (Gen 1:1). Accordingly, creation begins the entire biblical storyline. It sets the plot and trajectory of God’s entire plan. It undergirds the progression of scriptural revelation and theology. And because it is so foundational, it also formulates one’s worldview. Consequently, the opening chapters of Genesis set one’s perception of the sciences. That is not only because creation constructs the very structure of one’s worldview but also because it has direct bearing upon science itself. The opening chapters of Genesis account for the very origin of the material and phenomena that the sciences observe. For these reasons, theology classically has been known as the “queen of the sciences.” It is the overarching standard of truth and the very framework in which all the sciences are based upon, operate, and
However, questions and challenges have arisen concerning this passage of Scripture. The theory of evolution has provided an account of origins apart from any notion of Creator or creation. Evolutionists have pointed to numerous observations in support for their assertions [Mey17]. Due to the persuasiveness of these arguments, Christians have considered and even adopted aspects of evolution to varying degrees. On one end of the spectrum, there are some who entirely reject evolutionary accounts of origins and argue for young earth creationism. There are also others on the opposite end who argue for theistic evolution, which contends that God used evolutionary processes to formulate this world [Mey17, pp. 40–43]. And there are people in between these two extremes. Because of such significant controversy, some wonder if creation matters. It seems to be a contentious and unclear issue. Good people disagree. So some may believe that perhaps the Bible does not provide enough information to reach a definitive conclusion, and one should not have a dogmatic view on this. Moreover, people often perceive creation as something which took place so long ago that it has little pertinence upon the present time and upon more weighty doctrines in the Christian faith. For these reasons and more, the temptation is to isolate these passages in Genesis to something that is ambiguous and should not factor heavily in how we understand the world and science. To some, Genesis 1–2 can simply be relegated to a text that can have multiple interpretations. At that point, creation has become a pawn.

The goal of this article is to reclaim theology as the queen of the sciences and creation as a key starting point. This article will survey through the doctrine in its scriptural presuppositions, substance, theological stakes, and significance. In doing so, a growing case emerges that the doctrine of creation is not negotiable, secondary, isolated, irrelevant, or a liability. Rather, it carries divine authority and clarity, acts as a cohesive force in Christian theology, and even provides some of the most compelling answers to the most profound questions of life. Creation is not a doctrine to be isolated or set aside; it should not be an afterthought in the pursuit of science. Rather, the truth of creation, by virtue of its character, should be put on center stage. For it not only wields the definitive authority to determine our understanding of the sciences, but also the beauty to show the nobility of the sciences. Creation then is truly part of the queen of the sciences.

**Scriptural Presuppositions of Creation**

The entire discussion of creation and the centrality of theology must begin with bibliology. After all, the question of the queen of the sciences all depends upon how one perceives reason versus revelation. In recent history, the advent of modernism has entrenched a skewed view of the nature of Scripture and human understanding in popular thinking. This is all in the name of objectivity [Mor03]. Ironically, what is needed is to recover true objectivity about
the nature of revelation and the limits of human knowledge. Only then can one understand
why theology is the queen of the sciences. Such a framework is philosophically necessary
as revelation can alone provide the basis for and the authentication of what science depends
upon and assumes. In that way, bibliology is the cornerstone to the response of how creation
and theology relate with the sciences.

A helpful way to discuss this is to define and think through three categories: special rev-
elation, general revelation, and knowledge. Within each category, we can survey through the
source of each type of information, what it covers, how it covers it, and what it accomplishes.
Doing so will allow one to put revelation and reason in their proper places.

Special Revelation

With this in mind, the first category to cover is special revelation. The source of special
revelation is God Himself, for in it He directly reveals His truth (cf. Deut 29:29; Eph 3:5–9).
This can happen through a variety of means including direct acts of communication, dreams
and visions, and the incarnation of Christ who explained the Father (cf. Jhn 1:18) [MM17,
p. 74]. Special revelation also occurs in Scripture. In fact, Scripture is the foremost means
of such revelation for it is the repository of the revelatory acts God desired His people to
know (2 Tim 3:16b) [Rey98, pp. 116–26]. This is the point of verbal plenary inspiration. This
doctrine articulates that while men were involved in the writing of Scripture, their fallibility
never contaminated any aspect of Scripture. Rather, God superintended by His Spirit that
every word—all Scripture—is His own as men spoke from Him (Exod 4:12, 15; 2 Tim 3:16; 2
Pet 1:21) [Rey98, pp. 25–44]. Inspiration declares that Scripture is not in any way from man
but totally from God. It is pure special revelation.

The primacy of Scripture as special revelation is seen in that God Himself traces the work
of special revelation in terms of the giving of Scripture. Hebrews 1:1–2 notes, “God, having
spoken long ago to the fathers in the prophets in many portions and in many ways, in these
last days spoke to us in His Son.” These verses provide the breadth of God’s work of special
revelation. The descriptions of “long ago” and “in the last days” exhibit God’s revelatory
activity throughout all history [Bru64, Ell93, Lan91]. The verses also mention God’s “prophets”
who convey God’s direct message, “many ways” which include visions and dreams, and God
working in “His Son” which deals with Christ’s ministry. As noted, special revelation includes
all of these means. Nevertheless, they all take place “in many portions” of inspired writ
and revolve around God “speaking,” which refers to that which He verbally communicated.
Scripture is at the core of special revelation. In fact, this entire framework about all of God’s
activity of special revelation is given by Scripture itself—in this case, it is given in the book
of Hebrews. Accordingly, epistemically speaking, one only knows of the nature of special
revelation by what is revealed in Scripture. This illustrates that the ultimate foundation for
grasping special revelation is found in God’s Word. With that, God is the source of special revelation and Scripture is the foremost giving of that special revelation.

What God reveals in Scripture is staggering. This includes the theological depths of the various attributes of God in their loftiness (Isa 6:1–3; Ezek 1:1–2:1), transcendence (Isa 40–48; Job 26; Pss 96–99), and compassion (Exod 34:6–8; Ps 103). Such revelation also involves the nature of sin (Ps 14; Rom 1–3) and salvation (Isa 52:13–53:12; Rom 3:21–26), morality based upon His holiness (Lev 19:1), the work of the Spirit (Ezek 36:26–28), and the preeminence and glory of His Son (Rev 4–5). However, God does not merely reveal theological information in Scripture. Rather, the claims of Scripture also entail matters of history (1 Kgs 15:7) and observations about the phenomena of this world (Job 36:27–33). Indeed, these claims are a crucial part of the theological assertions of Scripture. God’s theological assertion about His plan for the world necessarily entails historical facts from beginning to end (Gen 1:1; Rev 22:1–9). Similarly, God’s theological assertion about the nature of man is predicated upon his origins. Thus, as will be shown, if the historical claims of Scripture fail, its theology collapses as well [Cho14, pp. 24–33].1 We cannot have one without the other. This is where the doctrine of inerrancy enters in. That doctrine reminds us how Scripture has no error in any of its assertions in the original manuscripts; all that it declares in every aspect of that declaration is true (cf. Jhn 17:17) [Bea06, Cho16]. Consistently, special revelation makes authoritative claims about reality that cannot be limited to just spiritual ideas but include a plethora of details of what takes place in this world. Even more, Scripture reminds us that its details are not merely without error but also disclose the full truth about these facts. Man may observe something but not know the purpose behind it (cf. Ecc 3:11). But revelation explains the data the way God wholly intended it to be (cf. Deut 29:29). Furthermore, Scripture contends that its information has far greater certainty than human experience. In Peter’s final epistle, he recounts how he was an eyewitness to Christ’s glory (2 Pet 1:16–18). However, although his experience was so compelling, the apostle proclaims, “And we have as more sure the prophetic word” (2 Pet 1:19). Peter declares that Scripture is even more sure than his experience of God’s revelatory activity. Scripture provides greater certainty than one’s own senses, memory, or observation. Consequently, special revelation covers a diverse set of information with a purity, depth, and certainty that goes beyond human derivation.

Scripture is also effective. As Psalm 19 says, God’s Word converts the soul, makes wise the simple, enlightens the eyes, and stands forever (Ps 19:7–9f). The Spirit uses His Word to transform God’s people. Second Timothy 3:16 not only affirms Scripture’s inspiration in that it is God-breathed, but also that it is “profitable for teaching, for reproof, for correction, for training in righteousness.” Scripture, under inspiration, contains the revelation of God that people for all time need to hear, communicated in perfection, and empowered to be effective

---

1 See also later discussion.
among the saints. With that, special revelation, and in particular Scripture, occupies a unique place in thinking about information. It covers a great breadth of data from theology to history, from the supernatural to the natural order, for it makes claims on all of this. It also has great depth, for it reveals with absolute certainty and truthfulness the full truth about life and the world categorically and in their details. Its information is also not inert. It has the power to save and change lives forever. Already, one might be able to observe why theology is the queen of the sciences. With such range and authority, it covers every discipline. And when something is done in accordance with God’s Word, it elevates the dignity of that activity.

General Revelation

That being said, there is a second category to be discussed, and that is the category of general revelation. Since this is a form of revelation, the source of this is from God. As Romans 1 describes, God Himself has made this revelation evident (Rom 1:19). Consistently, general revelation declares absolute, inerrant, and authoritative truth about God, truth that man could not otherwise determine. However, unlike special revelation, the means of this revelation is different. As opposed to being in Scripture, general revelation is often associated with creation. For example, Psalm 19 describes how the heavens proclaim the glory of God and how nothing is hidden from the heat of the sun in the sky (Ps 19:1–6). Likewise, in Romans 1, it talks about how God’s power and divinity are known through what has been made (Rom 1:20). However, there is a significant qualification about the means of creation as general revelation; namely, it is found in creation categorically as opposed to in individual details. Psalm 19 emphasizes the existence of the sky and the sun as a whole as opposed to a particular property of these entities. It stresses how all people are exposed to those elements. Similarly, Romans does not speak of individual components of creation but rather creation in its collectiveness—that which has been made (τοῖς ποιήμασιν). Accordingly, while general revelation is found in creation, it refers to creation in totality and not in its detail, as a whole and not in part. Such universality is not incidental to the category of general revelation. After all, a major purpose of general revelation is to make all without excuse, and that can only occur if general revelation is found systemically in all creation. If the means of revelation was located in a specific feature of creation or within man’s particular observation, not everyone may have access to that information or could figure that idea out. Because of this lack of access, man would then have an excuse before God. However, the entire point of general revelation is to show that all are aware and condemned. That requires the means of such revelation be found throughout creation. As the name general revelation suggests, the very means of conveying this revelation

---

2 Most likely, akin to the rest of the plurals in the verse, this is a form of collective plural [Wal96]. Paul also implies that conscience falls into general revelation. In Romans 2:14, he speaks of how conscience condemns people, proving that they are a law unto themselves [Gato2].
is general.

That is not the only way general revelation is general. It is also general in its content. Both Psalm 19 and Romans 1 make clear that general revelation pertains to the existence of God and His power. Psalm 19 speaks of the “glory of God” (Ps 19:1) and Romans speaks of God’s eternal power and divine nature. The latter does not refer to God’s various attributes but the quality of being divine (θειότης) [DB00]. In other words, with general revelation, people recognize that the concept of God is true but do not necessarily know who this God distinctively is. While this information is enough to condemn, making all men without excuse (cf. Rom 1:20), it is severely restricted. It only expresses a singular truth about God’s existence and not anything about other categories of theology, God’s biblical theological plan, and certainly not about His work at the beginning of that plan; namely, creation. General revelation is very general in that it covers a very limited set of truths in a very limited way.

Consistently, general revelation is general in its effect. As mentioned, it does not have the ability to save but only condemn. In addition, it can be inhibited. Paul begins an entire discussion of general revelation with the reality that men “suppress the truth in unrighteousness” (Rom 1:18). With that, general revelation, unlike special revelation, is truly general and thereby quite restricted in its means, content, and effect.

Knowledge

One third and final category remains: knowledge. The Scripture often discusses the area of human understanding (cf. Job 28:1–11; Prov 3:5–6; 1 Cor 1:18–31; Col 2:8; 1 Tim 6:20). Unlike divine revelation, the source of knowledge is man himself. To be sure, any true knowledge ultimately comes from God who established it in this world (Prov 8:22–31), but unlike revelation where God is directly revealing (cf. Rom 1:19), knowledge comes as man discovers and discerns. In that way, there is a categorical difference between revelation (whether special or general) and knowledge, and such a distinction results in other marked differences between the two. Unlike revelation, Scripture describes how knowledge can take truth and pervert it (Rom 1:21–23, 28). It is unsure as it assumes something but only under cross-examination discovers something different (Prov 18:17). It is erroneous as it discerns a certain path which leads to death (Prov 16:25). It can be futile as one only learns things that are incomplete and partial. That is why Ecclesiastes says, “Because in much wisdom there is much vexation, and whoever increases knowledge increases pain” (Ecc 1:18). Solomon states this in context because in learning facts, one cannot see all that God intended by them so one learns of problems without solutions, insights without balance, criticism without hope, and ideas without purpose [Enn11]. Those deficiencies of human knowledge lead to vexation [Lan92]. Human knowledge can not only be erroneous and limited but fallible. Knowledge may give one a sense of confidence, but Scripture observes how spiritually blind people grope in the dark
to find truth but cannot (cf. Acts 17:27; Job 12:25). While human knowledge can span the breadth and depth of creation, its understanding of these things is far from the certainty and completeness of revelation. That is not only because finite people are involved but even more so given man’s depravity (cf. Eph 4:17–19) and how they have an unfit mind (cf. Rom 1:28). Thus, any fact or aspect of truth that one can observe is often obscured by sinfulness.

Having defined these three categories, one can see that man’s temptation is to confuse them. At times, people observe that general revelation is tied with creation and assume that whatever they observe from creation then must be general revelation. Such an assumption gives rise to the adage “all truth is God’s truth” [Jon3]. However, that is incorrect. As noted, the mechanism of general revelation is not the individual components within creation but the very whole of creation itself. Moreover, its revelation is not discerned by man’s observation but by God’s revelatory activity, as He unveils it to each man (Rom 1:19). It is not man discovering things about God but God using the totality of creation to declare truth to man. Furthermore, the truth that is declared is not all the insights people think they might have from creation. Rather, it primarily concerns the truth of the existence of God. Thus, the notion that man can gain theological understanding solely by reasoning through aspects of creation is not what general revelation is nor how it works. What people have at this point is their knowledge and not revelation, and it is a category mistake to elevate their fallible reason over inerrant revelation. Similarly, people may also identify God’s character displayed in various details of creation, assume that these details themselves revealed such truth, and therefore that their discipline is engaged in revelation. However, that is not exactly what is taking place. Theologically speaking, when people observe the character of God manifest in His created order, the created order is not the source of that information. Scripture is where that revelation takes place. That is where people have learned with certainty about who God is. They then apply what they have learned to their discipline. This is an important reminder that knowledge is not inherently evil. It can and should be used for God’s glory; that is how God designed it to be (cf. Pss 8:3; 66:3; 92:5; 111:2; 104:24). That being the case, the created order becomes the illustration of special revelation but is not special revelation itself.

Man’s temptation is to elevate knowledge to the level of revelation, but they are not on the same level at all. They do not have the same source and thereby do not have the same certainty, value, content, completeness, power, or authority. That alone demonstrates that knowledge can never have the same influence as revelation. Furthermore, the very nature of knowledge shows a critical necessity and dependence upon revelation. For instance, science itself has certain assumptions including the reliability of one’s senses, the existence of an independently verifiable reality, and the presence of truth that can be discerned and articulated [Mor03]. However, postmodern thought has demonstrated significant philosophical challenges to those presuppositions [Mor03, Mor17]. How can one know that what he has seen or experienced
is true? Even if someone validates that experience, how can one even know that the person had the same experience, conveyed true affirmation, and even exists (as opposed to being a figment of one’s imagination)? These epistemological problems are significant. However, if such things cannot be definitively proven, then science—which relies upon observation, the existence of a universal reality, and the verification of repeatability—cannot work [Mor17]. This illustrates that man’s reasoning cannot autonomously deduce a self-sufficient system of thought. In actuality, the more they think, the more they realize, as wisdom literature long ago observed, that there are sizeable gaps in man’s assumptions and that this too is futility (Ecc 1:18; Job 9:1–2; 12:1–12; 20:1–16)3. For science to work, much must be assumed which cannot be definitively proven by human reasoning. Revelation is not just a crutch for the weak; it is a philosophically necessary crutch for science as well.

The book of Job illustrates the very limitations of human knowledge and the necessity of revelation. In fact, it is more than illustrative. As the chronologically first book of Scripture, its function is to introduce Scripture and, within this, to demonstrate why one needs a Bible to begin with [Wal12, Lon12]. That is why the book spends a considerable amount of length discussing wisdom. While suffering is found at the beginning of the book and drives the plotline, the vast majority of its chapters is consumed with dealing with the debate between Job and his friends. They endeavor to figure out why Job suffered. And they pull on the breadth of human knowledge to answer this question. They draw on history (Job 5:27), philosophy (Job 11:1–7), and frequently observations from nature (Job 8:11–18; 25:1–6). However, none of Job’s friends ever come close to the right answer. None of them ever articulates anything close to what took place in Job 1–2. In actuality, most of their conclusions blatantly contradict the truths stated therein (Job 8:20; cf. Job 1:1). Furthermore, what they do say, even if legitimate in certain circumstances, only leads Job to further despair. Therein is the nature of human knowledge. It can be brilliant. Some of what Job’s friends say are sublime descriptions of God and this world (cf. Job 11:1–20), even affirmed later by New Testament writers (1 Cor 3:19; cf. Job 5:13). Nevertheless, these facts lack completeness and so provide no comfort to Job. All of this leads Job to conclude that man is lost, groping in the dark (cf. Job 12:25). Human knowledge is not the panacea that one might think it is. Human observation can actually draw a completely wrong conclusion and make things worse rather than better. Having said this, one cannot fault Job’s friends too harshly. How could they know what truly and completely

3 For example, in Job 9, Job responds to Bildad’s scientific explanations with “In truth, I know this is so.” Man does make correct observations about this world and its operations. But it cannot solve the question “But how can a man be right before God?” Science cannot see everything. This is also brought out later on in Job 38 where God answers Job with a challenge of where he was when God laid the foundations of the earth or if he ever commanded the morning. Such statements remind Job of his inherent limitations in making observations and deducing God’s character from this world. Job simply has not seen enough of the entire picture to make such a judgment call. See also [Enn11]: “The reason why even his royal wisdom quest is futile is that the more you know, the worse off you are. It is futile not because of what he cannot find out, but because of what he does find out. As he puts it, wisdom is accompanied by sorrow, knowledge by pain.”
took place? That involves heaven itself (Job 1:6) and thus only heaven could reveal it. And that is precisely why revelation is needed. Human knowledge cannot know things with certainty or completeness. Therefore, heaven must disclose it.

Toward the end of the dialogues, Job reflects upon all of this. In Job 28 he begins by acknowledging that man has tremendous talent and intelligence. His advancements in mining illustrate this (Job 28:1–11). Nevertheless, he questions where wisdom is truly found and concludes that man has no ability in and of himself to perceive (Job 28:12–14), acquire (Job 28:15–19), or invent wisdom (Job 28:20–22) [Har88, Ald93, Wil15]. The reasons are simple. Man, being part of the world, cannot stand outside of the world and objectively observe the complete picture of what is taking place (Job 28:12–13). Man also does not have the power to control the world and thereby his knowledge does not wield such effectiveness. He only has money, which is a poor substitute for true might and wisdom (Job 28:15–19). In addition, man does not have the skills to discern wisdom, for it requires the ability to know all things natural and supernatural (Job 28:20–22). Job acknowledges that man’s knowledge can be astounding, but he does not have complete wisdom. And in context, this is Job’s realization of why he and his friends cannot figure out his problems. Instead, Job realizes that the solution is found elsewhere. As he says in the final verse of Job 28, “So He said to man, ‘Behold, the fear of the Lord, that is wisdom; and to turn away from evil is understanding.’” Why is fear of the Lord the beginning of wisdom? God alone is the one who sits outside creation and sees all things (Job 28:24). God alone is the one who controls creation and has the knowledge that has true power (Job 28:25–26). God alone is the one who knows all things natural and supernatural and has discerned all things (Job 28:27). Therefore, there is one who knows what He is talking about, and there is man whose knowledge is shallow, incomplete, and errant. When one fears God, he listens and surrenders to Him. And in that moment, man finally becomes wise because he heeds the only one who knows what He is talking about. The book of Job is a lesson on the strict limits of man’s understanding and that if one is to have any real wisdom or answers, one must fear and surrender to God. Without this, one will appear smart like Job’s friends but will also be just as foolish and unhelpful as they were. Man needs revelation to figure out life.

As stated above, out of any line of argumentation, bibliology is what most definitively locks in theology as queen of the sciences. That is because it demonstrates the distinction between reason and revelation. It shows how revelation is more certain and complete than reason, how revelation undergirds all of reason, and how reason then can never trump revelation. As Job has observed, man on his own cannot come up with the answers, God must supply them, and thus the posture of man cannot be one of intellectual self-reliance but of the fear of God. With that, theology acts as the starting point for the sciences, not only relative to authority and sureness but even in philosophically accounting for the assumptions of the discipline. For
in Scripture are the reasons why we know that reality, truth, the reliability of our senses, and even communication all exist. Philosophers have a point. We may not be able to justify these things on our own, but we do not have to. God has revealed it so, and the grounding for that is in creation—the way God made this reality, established it in wisdom and truth (Prov 8:22–31), and made man able to understand and communicate (Exod 4:11). With that, theology and even creation itself are queen of the sciences.

The Substance of the Doctrine of Creation

At this point, one might acknowledge that theology occupies an authoritative place over science. However, an issue arises of whether we can really know what Genesis 1–3 says. How can we be sure that Genesis 1–3 reads as complete history of what took place at the beginning of the world as opposed to having some flexibilities that can incorporate what people have observed in evolution? If creation cannot be confidently defined, then its place as the queen of the sciences is simply a theoretical point. An undefinable definition cannot define anything.

Hence, the question is: What does Scripture say? Is there a way to discern the Bible’s claim about the nature of the opening chapters of Genesis? I have commented on this in other publications and will summarize the line of argumentation here [Cho14, pp. 19–46]. Essentially, three major hermeneutical options exist for thinking through the claims of Genesis. First, one could argue that parts of the Genesis account are not actually claiming to be historical, which allows for a certain level of accommodation to evolutionary claims. Second, one could argue that the whole of the Genesis account is not claiming to be historical which would then provide no obstacle in affirming an evolutionary perspective to whatever degree one desires. Third, Genesis may be claiming to be historical both in part and whole. If the latter is the case, Scripture is incompatible with an evolutionary conceptualization of origins. In essence, discerning this issue is a process of elimination. By disproving the first two options, we establish the third.

Evaluating Whether Particular Claims are Historical

With this in mind, one can broach the first hermeneutical possibility. Does Genesis 1–3 contain certain details that do not claim to be historical? Examples of such thinking would include gap theory, day age theory, Genesis 1–2 as poetic genre, phenomenological language, and the need for harmonization.⁴ While each of these ideas comes at the subject from a unique angle, the end result is the same. They contend that certain descriptions in Genesis are ambiguous

⁴See discussion in [Roo92, pp. 411–27] and [Kli58, Kid67].
and not meant to be read strictly as history. There is flexibility in the text that does not put it in conflict with what man has observed.

However, these theories do not stand scrutiny. For example, gap theory argues that a disjunctive clause with the perfect tense of the verb “to be” in Hebrew implies that Genesis 1:1–2 is not necessarily contiguous and may even imply that there was a deconstruction of original creation as the earth became empty and void. However, an investigation of both the disjunctive clause as well as how the perfect verb היהדרו operates does not support such conclusions [Roo92, pp. 316–23, 411–27]. In fact, even a standard Hebrew grammar defines disjunctive clauses as often operating as circumstantial clauses delineating the very situation of the main clause [WO90]. The syntax of the text is not ambiguous or suggestive of what gap theory alleges at all. It is clear in its presentation.

Likewise, a day is also defined in Genesis 1. In Genesis 1:5, Moses actually uses a cardinal number (ָאִלּוֹד, one) as opposed to an ordinal number (ָאִדַּל, first). All the rest of the days are numbered by ordinal numbers, so the initial exception is significant. In saying, “And there was evening and there was morning, one day” (Gen 1:5), the narrative does not merely talk about the first day but defines what constitutes a single day [Ste02]. It is evening and morning. The text itself defines its own terms contrary to day age theory.

Similarly, assertions that the opening chapters of Genesis are poetic are also linguistically tenuous. A major marker of historical narrative is the presence of the wayyiqtol verb form [Boy08, pp. 163–92]. Genesis 1–3 statistically has this verb form in the exact ratio expected for narrative and not poetry [Boy08, pp. 163–70]. To argue that Genesis is poetry and thus has a more literary flexibility in interpretation is baseless. In fact, such an argument is additionally flawed in that poetry can convey historical facts (cf. Pss 78; 105–6). Genre does not determine the content of a text but rather sets the presentation of those claims and the desired response [Tho02].

Another suggestion has been that since Genesis (or really Scripture as a whole) contains phenomenological language or language of appearance, it must not be historical [Lam13, p. 50]. To be sure, stories throughout Scripture record things from certain people’s perspective. A famous example is how Scripture records how the sun rises and sets (Gen 19:23). However, such claims do not undermine the accuracy or historicity of a text. For one, the text claims that phenomenological language is what one observes from his perspective. Such an assertion is true and accurate. Furthermore, for such an observation to occur, it presumes one saw something which places the experience in the realm of time and space. Thus, even though an event may be redescribed from an external perspective, such a truth does not negate that an event exists or demand a parabolic interpretation.

Finally, some argue that potential conflicts between Genesis 1 and 2 argue against a strict historical reading of the text. For example, Genesis 1:11–12 says that vegetation sprouted, but Genesis 2:5 says that plants had not yet grown. Another supposed conflict occurs when Gen-
Genesis 2 talks about how God formed Adam (Gen 2:7) and then formed every beast of the field and bird of the sky (Gen 2:19). However, Genesis 1:20–24 states that the creation of animals occurred before Adam was created. Because of these alleged contradictions, some argue that Genesis 1 and 2 cannot be read as precise historical narratives since the accounts do not agree on what happened. Instead, they contend that a more generic reading is preferred which then allows for Genesis 1 and 2 to harmonize with each other as well as with current scientific thinking [Wal13, Futg8]. However, the supposed inconsistencies can easily be reconciled. The plants mentioned in Genesis 1:11–12 are not the same terms or types as mentioned in Genesis 2:5. Trees and grass, which are not cultivated by human beings, were created fully grown as Genesis 1 explains. But cultivatable plants, like shrubs or plants of the field, were created in their seed form awaiting man to nurture them as Genesis 2 states. It is not a contradiction at all if different plants were created in different ways as Genesis 1–2 describes. The alleged conflict concerning the ordering of the creation of animals and man can be solved similarly. The animals described in Genesis 1 are also not the same terms or types of those found in Genesis 2.5 Furthermore, for this particular issue, another point of explanation is to understand that the verb form in Genesis 2:19 does not always have to deal with strict chronological sequence, especially when the clause provides the explanation of the circumstance surrounding a situation (Num 1:47–49; Ruth 2:3; 2 Sam 14:5; 1 Kgs 13:12). That is precisely the case here [Ham90, p. 176], [Mat96, p. 214]. The text may very well be saying that God had already formed these animals to bring to Adam to name. This understanding is reflected in quite a few modern translations. In light of this grammatical reading, both Genesis 1 and 2 provide the same order after all.6 The supposed conflicts between texts are hardly insurmountable.

Does Genesis 1–2 have certain details that are ambiguous or interpretatively flexible? The above analysis has demonstrated that this is not the case. In fact, it demonstrates the opposite. Genesis 1–2 not only have clear grammar and defined terminology, they even employ very technical terminology, as the designation of plants and animals in these chapters demonstrates.

5 Genesis 1 does not mention the creation of the “birds of the sky” (םוֹנֶ הַתְּלוּ) which is mentioned in Genesis 2. Likewise, Genesis 1 only mentions the “beasts of the earth” (נִ הַם הָאֵרֶץ) whereas Genesis 2 mentions the “beasts of the field” (נִ הַמֶּלְאךְ הַגּוֹיִם). The animals mentioned in Genesis 1 are all the land and sky animals that roam freely or are livestock, but the ones in Genesis 2 are land animals that are used to cultivate the land and birds that man has dominion over. In fact, God even mentions in chapter 1 that the “birds of the sky” are those birds which man has dominion over (cf. Gen 1:26).

6 To be clear, both factors of the distinction between animals named in Genesis 1 and 2, and the fact that Genesis 2:19 must be reflecting on what God had already done before Adam’s creation are not mutually exclusive. It could very well be that God formed these unique sets of animals before He made man and then later brought specific animals to him to name. That best accounts for all the evidence presented thus far [Ham90, p. 176], [Cas61]. The two supposed sets of discrepancies given above actually share a parallel. All plant life was created on the day given in Genesis 1 but different plants were created in different ways to serve different functions expressed in Genesis 2. In the same way, a pluperfect reading of the verb and acknowledgement of the technical designation of animals also recognizes that all animal life was created on the respective days given in Genesis 1 but with different classifications and emphases to be utilized for a purpose in Genesis 2.
All of this argues that Genesis 1–2 is a carefully thought out and crafted narrative, one with attention to detail. Such exacting precision in the narrative argues against the notion that there is hermeneutical flexibility within it.

Evaluating Whether the General Tenor of Genesis 1–2 is Historical

In the history of interpretation, theories like gap or day age have been abandoned by scholars in favor of alternative approaches that argue that Genesis 1–2 is designed more like ancient near eastern (ANE) myth [Wal09]. This shifts the discussion from discerning whether particular claims of Genesis are historical to whether the entire account intends to be historical. If Genesis 1–2 corresponds more to the myth of Israel’s contemporaries, perhaps the text is more like a parable, intending spiritual truth but not necessarily telling a historical story.

The support for this suggestion comes from the extensive overlap between Genesis and ANE myths. Some of the most dominant comparison are as follows:

- The Hebrew word “deep” in Genesis 1:2 sounds like the word for the goddess Tiamat of the Ancient Near Eastern creation myths.
- Both have waters above and beneath separated by a firmament.
- Both have light before the sun, moon, and stars.
- Both describe mankind’s failure to please a deity.
- Both refer to plants that confer immortality.
- Both mention a serpent.
- Both describe a global flood [Cho14, p. 34],[Lam92, p. 527],[Gro85],[HY00, p. 451].

At first glance, these seem quite compelling. However, two points should be raised. First, on a deductive and theological level, Scripture maintains itself as the truth and does not accommodate falsehood. Paul repeatedly repudiates myths (1 Tim 1:4) and exhorts any leader in the church to do the same (1 Tim 4:7; Tit 1:14). Heeding myths is the sign of a false teacher (2 Tim 4:4). One may object that the myths Paul talked about were not the stories of the ancient near east; nevertheless, the rhetoric points out that Paul’s objection against the false teachers is that they taught things that were a tale or legend, that which was not true but fictional. While targeting a specific issue, he deals with the entire category [Bea06, pp. 301–33]. Consistently, Peter asserts that in presenting the life of Christ, they did not follow myths but were eyewitnesses (2 Pet 1:16). The contrast between myths and an eyewitness account is one of fiction versus fact, fantasy versus verifiable history [Bau83],[Sch03, p. 313]. The notion of affirming incorporation of legendary material into Scripture is problematic from the claims of the biblical writers who composed it.
Second, inductive analysis supports what the biblical writers assert. In comparing ANE myths with the Genesis account, the similarities are not as similar as one might perceive. The list above is compelling, but the list above is the result of an interpretative process. For instance, scholars have observed that both Genesis and ANE myths have waters above and beneath separated by a firmament. While Genesis states this directly (cf. Gen 1:7), ANE myths speak of how one deity split a goddess in half and put half her body above the sky and the other half below it [Lam92, pp. 135–40]. Such a description is far from what Genesis articulates. Genesis talks about physical matter; the ANE myths speak of deity. Genesis talks about God’s creation; ANE myths speak of battle. Genesis speaks of a singular God making; ANE myths speak of a pantheon in conflict. In essence, as scholars note, Genesis is different than ANE myths in that Genesis demythologizes these stories [Wat27]. In using such a term, these scholars acknowledge that Genesis is far from myth. Consequently, there are substantial differences between Genesis and myths. On top of what has been just mentioned, Genesis does not use sexual imagery to discuss creation [HY00, p. 6]. As mentioned, it does not use poetry which is a main vehicle for ANE myths. It also does not use terms of evolutionary development, which ironically are present in certain myths of the time. For example, Egypt has the term “evolve” which lexically seems to denote such a process because it is used to describe how a chicken develops from an egg. Furthermore, Genesis uses temporal markers to discuss a distinct cosmology as opposed to ANE myths which are more cyclical in nature and describe the seasons [Gro85]. Put differently, Moses deliberately avoids the terminology, genre, temporal markers, descriptions, and religion associated with ANE myths. While modern scholars may be drawn to make comparisons, those in the ancient times would have seen these stark distinctions and understood that Genesis was anything but ANE myth. Therefore, the idea that the biblical creation account posits itself like a parable or fictional story is not sustained both by the nature of Scripture in totality as well as by the specific assertions in Genesis 1–3. Genesis presents itself as history as opposed to ANE myth.

Evaluating Whether Genesis Claims to be Completely Historical

By process of elimination, if Genesis does not make ahistorical claims in individual details or general tenor, then it is historical in both part and whole. It is completely historical. Two lines of evidence support that Genesis 1–3 asserts itself as history. Fundamentally, the mentality of Scripture as a whole ties theology with the reality of history. A simple proof of this is found in Paul’s logic concerning the resurrection. In 1 Corinthians 15, the apostle does not say that our Lord’s historical resurrection is irrelevant as long as the theology of the resurrection can be sustained. Instead, he contends that if Christ did not rise from the dead, then the theology

---

7 See [Cho14, p. 36, fn. 40].
of the resurrection in its hope and power is non-existent (cf. 1 Cor 15:1–19). Theological truth is tied with history. However, the rationale that history and theology are tied together extends beyond the resurrection. For example, Romans 5:8 ties the historical reality of Christ’s death to the demonstration of God’s love. Similarly, Peter ties the global flood to the reality that God will judge the earth once again (2 Pet 3:4–6) [Scho3, p. 374]. In addition, Hebrews 11 ties the historicity of Old Testament events to the efficacy of faith. In that specific passage, if the events listed did not happen, how can one argue that faith is effective when it never really had an effect? Interestingly enough, Hebrews 11 includes creation in the list of history it recounts (cf. Heb 11:3). The resurrection is not the only event where its historicity is mandated for theological reality. The logic found in 1 Corinthians 15 is truly found throughout the argumentation of Scripture. Accordingly, one expects this pattern when reading Genesis, especially since, as argued above, it makes no claim to the contrary.

In addition, the way Genesis presents itself is the way Scripture reads every part of it. Moses himself constructs subsequent narratives in the Pentateuch to uphold a historical reading of Genesis 1-3. Even in Genesis 4, Adam is treated as a real human individual who has relations with his wife and has children (Gen 4:1). In the giving of the Ten Commandments, Moses records how God Himself declares, “For in six days Yahweh made the heavens and the earth, the sea and all that is in them, and rested on the seventh day; therefore Yahweh blessed the sabbath day and made it holy” (Exod 20:11). The directness of this quote is significant. The Lord could have used comparisons, analogies, or even language of similarity all of which are found in the Pentateuch about other matters (cf. Gen 3:5; 13:10; Exod 15:11; 24:17). However, that is not the case. The quote above makes an emphatic claim, without any qualification, about how God made the world in six days and rested on the seventh. With that, the Pentateuch affirms the historicity of the creation account. Other books follow suit. First Chronicles begins with the genealogy of Israel with Adam, again presuming his existence (1 Chron 1:1). Psalm 104, in poetry no less, recounts God’s wisdom in how He created the world, and the poem is structured around the order of creation provided in Genesis 1 [Kid]. Amos 4:13 speaks of God making the mountains and, in the next chapter, the stars (Amos 5:8). In addition to all of this, the Lord Himself affirms the reality of the Genesis account. Matthew 19 says, “And He answered and said, ‘Have you not read that He who created them from the beginning made them male and female, and said, “For this reason a man shall leave his father and mother and be joined to his wife, and the two shall become one flesh”?’” Jesus specifically affirms that God is the one “who created” Adam and his wife. This is a special creation of man as Genesis describes. Paul affirms the existence of Adam in paralleling Adam and Christ (Rom 5:12; 1 Cor 15:45). Luke does so as well in his genealogy (cf. Luke 3:38). Both Peter and the author of Hebrews not only affirm that God created the world but specific details within Genesis, including that He formed the earth from the water (2 Pet 3:5; cf. Gen 1:2, 7, 9) and that He created by His Word (Heb 1:2; cf. Gen 1:3). Paul likewise affirms
that God created light (2 Cor 4:6). This list is far from comprehensive, and one could also cite examples of how the historicity of the creation account in its details is affirmed by Daniel, Ezekiel, Isaiah, and other New Testament writers. Suffice it to say that no other reading of Genesis is presented in Scripture. Thus, to posit an alternative interpretation of Genesis is not merely contained to the passage but will have a ripple effect to a whole host of other passages as well. This is not an isolated interpretation.

Therefore, Genesis 1–3 does not have hermeneutical ambiguity. Its details are not ambiguous since its terms (like day and the grammar of Genesis 1:1–3) are clearly defined. Its tenor is not ambiguous as Moses goes out of his way to differentiate his writings from ANE myths. All of this sets up for Genesis clearly asserting itself as history, an assertion affirmed without reservation by the rest of Scripture. The way Genesis presents itself is the way Scripture reads it and there is no other interpretation. With that, there is a singular hermeneutical mindset about this passage in Scripture. And that clarity punctuates why the doctrine of creation matters. God wanted this truth understood. This is not a doctrine that is hidden (cf. Deut 29:29) or even one that has components that await further elucidation by later revelation (cf. 1 Pet 1:10–12). Rather, this doctrine was originally presented in no uncertain terms, and the rest of Scripture maintains such clarity by reiterating the same ideas repeatedly without modification. Scripture then does not view its teaching on origins as a mystery or something to be nuanced or qualified. Instead, it views this doctrine as so clear that the rest of Scripture assumes the reader understands it. Consequently, one cannot avoid it or neuter its forcefulness by appealing to its supposed ambiguity. This is a doctrine that is to be known in the way that all Scripture knows it. It is definable and thereby definitive.

What is at Stake in the Doctrine of Creation?

At this point, one can see why theology is queen of the sciences and that the doctrine of creation is clear. A lingering question still might be whether the doctrine carries great importance. One may allege that though the scriptural assertions of creation are clear, some flexibility should be offered because creation is not an essential doctrine. In some people’s minds, because creation is not that crucial, varying viewpoints on this view offer little danger to the faith. Put differently, is there really anything at stake with the doctrine of creation?

To be sure, pastorally, one should have great patience with people as they grow into the truth (cf. 2 Tim 4:2). This is not just for the doctrine of creation but for any doctrine or addressing any sin. Nevertheless, that is a completely different issue than whether the doctrine of creation is crucial. As noted, creation is at the beginning of God’s story of this world and the genesis of His revelation of theology. Change something at the beginning, and the ripple

---

8 See discussion in [Lac01], [Choi13, pp. 83–90].
effect is found throughout the rest. Replace the framework of creation with a different one (like evolution) and if one holds it consistently, a reworking of theology takes place. That is what this particular discussion endeavors to demonstrate. Using the ten major categories of systematic theology, one can observe that the distinctive activity of creation described in Genesis undergirds the breadth of Christian doctrine. Creation is far from an isolated truth. It is at the beginning and thereby at the foundation of it all.

Theology Proper, Christology, and Pneumatology

For instance, one could begin by examining the doctrines of theology proper as well as Christology and Pneumatology. One aspect of God’s distinctiveness is the fact that He creates. Scripture assigns God the unique role of Creator and everything else as created. This is not only God the Father but includes the second person of the Trinity and the Spirit. The Spirit hovers over the waters (Gen 1:2); He affects creation directly. And when Scripture desires to demonstrate the deity of Christ, it appeals to His role in creation (Jhn 1:1-2; Col 1:16). Within this, Scripture defines the manner in which God creates. The Hebrew word for create (ברא) is used uniquely with God as subject and is not the same term as “make” (עשה) or “form” (יצר) which can be used with human subjects [Sch97]. While “making” and “forming” stress how one makes a product from already existing material, the verb “create” emphasizes the personal and direct intervention of God to produce that which is distinctively other than what had existed [Sch97]. For example, David desires the Lord to create in him a clean heart (Ps 51:10), which is categorically different than what was present in the psalmist. Likewise, God creates praise in people who were rebellious; He produces a work in their lives that was otherwise non-existent (Isa 57:19). Creation is the exclusive act of God, and such involved and creative acts do not just extend to the creation of the cosmos ex nihilo,9 but also to particular things including animals (Gen 1:21) and people (Gen 1:27) which all specifically involve the term ברא. The logic of Scripture is that because God created in such direct fashion, He is directly over His creation (cf. Isa 40:26, 28; 42:5; 45:12). He claims total and pure authority over all He has created because nothing else can receive credit for the existence of that entity [Sch97, Smi09, Osw98]. However, if evolution is true, then a different set of events exist than what the Genesis account describes. At minimum, intermediary causes were involved in the generation of animals or humans or the entire world. If that is the case, then the very distinctiveness, otherness, and authority of God must be redefined.

One can add a further note on the ramifications of changing the doctrine of creation on the doctrine of God. It not only shifts His distinctiveness but also His character. If one attempts to harmonize evolution with Genesis 1–3, then death happened before the Fall. After all, death

---

9 Theologians argue for creation of the world from nothing is rightly based upon how the word ברא is used in the context of Genesis 1-2 [Sch97], [Mat96, p. 129].
is an inherent part of the evolutionary mechanisms that generated this world. The Fall only takes place after everyone is made. Accordingly, God declares that a world with death is good and very good (Gen 1:31). Death itself then is something that is good. That changes the very definition of God’s goodness. A shift in one’s understanding of creation consistently shifts one’s viewpoint on God’s (in all three persons of the Godhead) otherness as well as the definition of His attributes.

Anthropology

If the biblical account of creation is not historical, then the category of anthropology must change. For one, the distinction between man and animal is blurred. In Genesis, man is created in the image of God and in an entirely different manner than animals (Gen 2:7). This elevates man over creation (cf. Gen 1:28–30) to have dominion over it. However, if evolutionary thought is true, then man originates from animals and the distinction between people and animals is not as sharp, if present at all. Likewise, biblical anthropology argues that there are two biological sexes for man, male and female. That too is based upon the creation narrative (Gen 1:27). In fact, later discussions of sexual morality will use terms from the creation account to anchor ethics in what God did in creation. However, if the account of creation did not occur, then the idea of male and female has no actual grounding and as a result, biblical ethics about human sexuality come undone. In addition, the unity of the humanity is predicated upon creation. The idea that man is one is because people all descend from one father, Adam (and later on Noah; cf. Acts 17:26) [Boc07]. However, if this is not historical, then that unity has no true basis. The truth that people are ultimately of one family and should not act in partiality against one another has no foundation in reality. If one changes the origins of man, one changes a great deal of the category of anthropology and its ethical demands.

Ecclesiology

If a shift in the doctrine of creation changes the doctrine of man, it also will equally affect the doctrine of ecclesiology or the church. A critical component of the church is that the church is a new humanity in Christ (cf. Eph 2:15) [Thi10, p. 170]. The significance of the unity within the church demonstrates that the gospel can overcome the effects of sin and the fall, the most obvious being how man is fractured and divided though actually one. However, if the evolutionary account is true, humanity does not necessarily originate from a single head.

---

10 For example, Romans 1 does not use the common phrases ἀνθρώπος or γυνή but rather ἄρσην and θῆλυ (cf. Rom 1:26–27) which correspond to the Greek translation of Genesis 1:26-28. The same occurs in Galatians 3:28 which distinctively borrows from the exact wording of that passage (οὐχ ἐνί ἄρσεν καὶ θῆλυ). In the Old Testament itself, Leviticus 18:22 does not use the more common root עת for man but rather יְלִיו which goes back to Genesis 1:27.
11 See later discussion for further explanation.
evolution is at all true, the church demonstrates a unity humanity never had nor was intended to have. If evolution is at all true, a major theological significance of the church falls apart.

Hamartiology

A shift in the doctrine of creation would also affect hamartiology or the doctrine of sin. For one, as mentioned, creation is what grounds why particular sins are wrong (cf. Lev 18:22; Rom 1:26–27). A change in the doctrine of creation would remove the ground from certain ethical imperatives and thus alter the doctrine of sin. Moreover, one’s categorization of good and evil also should be modified if the creation narrative is reinterpreted. Scripture, since the beginning, has associated death with sin and evil (cf. Gen 2:17). It is something to be overcome as opposed to something that should be promoted (Heb 2:14–15). However, if an evolutionary backstory is adopted, death came before the Fall and is actually “good” in that it facilitates natural selection. If one changes the foundation of creation, the definitions of what is good and evil necessarily shift as we begin calling good what the Bible labels as evil. On top of this, the very scope and existence of evil are brought into question. The creation account provides the exact method of how sin came into the world (Gen 3:1–7) and shows how it affected the whole world (Gen 3:8–19). This thought presumes the existence of a perfect world, Adam, a tree of the knowledge of good and evil, Adam’s headship over humanity, and his disobedience. But if, per an evolutionary mindset, these things are myth and something else took place, then there is no true accounting for why evil is in this world, or if sin is universal among mankind or creation, or if it exists at all. In fact, given how death would have to be recategorized morally, perhaps good and evil do not truly exist as absolute moral categories with an evolutionary grounding of origins. Accordingly, sin both in particular and in category would be redefined if the doctrine of creation changes.

Soteriology

The doctrine of salvation would also be affected by a change in creation. After all, salvation is depicted as an act of creation. David cries out for God to create in him a new heart (cf. Ps 51:10). However, if “creation” is really just “evolution,” then what does salvation exactly do and produce in one’s life? Similarly, believers are called a new creation in Christ (cf. 2 Cor 5:17). However, if creation never truly occurred as Scripture describes, then what does a new creation denote? Furthermore, if, in any derivative of an evolutionary model, origins already included things like death, what does being a new creation mean? The very metaphors of salvation are distorted and undermined if the doctrine of creation shifts. These are not just theoretical questions since Paul himself discusses God’s act of creating light from nothing to show how God creates spiritual light in the darkened heart (cf. 2 Cor 4:6). Likewise, Paul parallels condemnation in Adam with salvation in Christ (Rom 5). In these contexts, Paul’s
logic sets creation and the act of salvation in parity. One is based upon the other. Accordingly, if creation did not historically occur the way Paul and Scripture describe, then the very logic and operation of salvation falls apart. This is of no small consequence to the category of soteriology.

**Eschatology**

Changes at the beginning alter the end. This is supported by the very language of new creation along with the language of a new heavens and earth (cf. Isa 65:17; 2 Cor 5:17; Rev 21:1), all of which parrot the original creation of heaven and earth in Genesis 1:1. Cosmology is tied with eschatology. However, if God never technically created to begin with, the definition of creation changes, and the definition of a new creation will change as well. If historically God originally created the paradise of Eden with death in a good world, why would one expect any different in the final creation? One might argue that the new heavens and earth are different because the Bible says so (Rev 21:4). However, harmonization with evolution raises serious questions to such descriptions. After all, if Scripture established that original creation was without death even though that was not the case, what grounds does one have to say any different about the end, especially when Scripture explicitly relates the two? With that, the very core of Christian hope of how God deals with sin and death is shaken if one consistently upholds a change in the beginning.

**Angelology**

There is also the category of angelology. One might think that an alternative understanding of creation would have little impact on this area of theology. However, in Job 38:7, God Himself recounts that the angels beheld and rejoiced over how He set the formation of the dry land (Job 38:3–6). If God did not directly create as Genesis specifies or if an evolutionary history is what actually happened, then this specific fact about angels is false. Even the doctrine of angelology is affected by the doctrine of creation.

**Bibliology**

The tenth and final category remaining is bibliology and to be sure, one’s position on creation and its potential harmonization with evolution will affect this. In light of redrafting the major categories of theology and all the passages therein, this is a redefining of the entire Bible, a matter of bibliology for sure. And that brings up what was discussed prior; namely, this entire discussion ultimately pertains to a discussion on how one understands biblical inspiration, inerrancy, and authority, as well as how these doctrines interact with accommodation and human observation [Lam13, pp. 48–53],[Gru17]. They deal with the nature of truth and
the hierarchy of certainty and truthfulness between special revelation and what one derives from “natural theology.” Such questions are a matter of the very substance of the category of bibliology, and so the way one answers the question of evolution and creation has bearing upon this area of theology [Gru17, p. 821].

The above discussion illustrates how each major category of systematic theology shifts if one incorporates evolution into the creation narrative and remains consistent with that. To be clear, this is not an allegation that those who contend for theistic evolution or other derivatives of accommodation or harmonization actually embrace these ramifications (although some do).12 Rather, the point is that logical consistency demands these consequences. And such consequences radically shift the major tenets that summarize Christian dogma, that which distinguishes Christianity as Christianity theologically. Thus, what is at stake with the doctrine of creation is nothing short of Christianity itself. And that demonstrates why the doctrine of creation matters. It is not an isolated doctrine that has little effect on the rest of Christian teaching.13 Rather, it is a doctrine that is at the beginning, thereby fundamental, and interwoven with the foundation of the chief aspects of Christian thinking. The doctrine of creation is one that cannot be ignored. If theology is the queen of the sciences, then creation is inextricably linked with that.

The Significance of Creation

The doctrine of creation has the backing of divine authority, clarity from God’s revelation, and is inextricably linked to the theology of Scripture. However, is there any significance to that? If it is part of the queen of the sciences, does it offer any value to the sciences? The following discussion shifts from what is at stake theologically concerning creation to the significance it has. There is a reason that creation is so entrenched hermeneutically and theologically in Scripture. It is because the doctrine of creation drives the very redemptive agenda of God in the storyline of Scripture. In essence, one could sum up the forcefulness of the doctrine this way: creation teaches that this is our Father’s world, everything in it is for His glory, and so He will make all things right in the end. That theology is not only a dominant driving logic in redemptive history, but is also at the heart of the answer to the major questions people have about this world. It thereby provides the underpinning logic for the beauty and telos of the sciences. The significance of the doctrine of creation is weighty.

12 See [Lam13, p. 58]. Lamoureux states, “Obviously, this conclusion challenges the traditional doctrine of original sin. Yet this is unsurprising, since it was formulated by anti-evolutionist and scientific concordists such as church father Augustine (364-430)” (footnote 32). Some in fact do hold to the ramifications listed above.

13 See [Gru17, p. 821]. Note Grudem’s statement: “Theistic evolution is not at all a harmless ‘alternative opinion’ about creation but will lead to a progressive erosion....”
An effective way to observe all of this is to survey from Genesis to Revelation. Within this, we can show how various points of redemptive history tie back to creation and display how creation drives God’s plan in resolving all things.

In tracing this, it is appropriate to begin at the beginning. And the creation account itself teaches the fundamental concept that this is our Father’s world, everything in it is for His glory, and He will make all things right. Each of these ideas is found within the opening chapters of Genesis. Fundamentally, Genesis 1 declares that this is God’s world. The first chapter of Genesis is structured around the different locales of light and dark (Gen 1:3–5), sky and sea (Gen 1:6–8), and the land (Gen 1:9–13). It then progresses to describe how God fills each of these locales: sun, moon, and stars for the light and dark; animals in the sky and sea; and then animals along with man for the land [Mat96, p. 144]. The very structure of Genesis 1 shows that God controls every place and everything that fills those places. This is our Father’s world. Furthermore, the text demonstrates that all is for His glory. God says that this creation is not just good but very good. God also sets the world apart for himself on the Sabbath day as He makes it holy unto Himself (Gen 2:3) [Ham90, p. 143]. With that, the whole world is consecrated for His honor. On top of all this, the initial chapters of Genesis teach that because He is over this world and it is for Him, He will make all things right. In Genesis 3, the Fall occurs but God has always had a plan. In Genesis 3:15 God specifically reveals that His Son, the Messiah, will ultimately crush the serpent’s head. This begins God’s agenda of redemption. Accordingly, Genesis 1–3 collectively establish that this is our Father’s world, everything in it is for His glory, and He will make all things right in the end.

This agenda is not merely declared at the beginning of the story but carried out on a global scale. The worldwide Flood demonstrates this. To be sure, the event exhibits God’s wrath in global destruction. At the same time, it is pertinent to observe that there are several parallels between Genesis 8 and creation. For example, in Genesis 8:1, the waters ascend above the mountains and to the point such that there is only sky and sea. The scene is reminiscent of creation (cf. Gen 1:2). In Genesis 8, there is a “wind” over the waters which sounds like the language of Genesis 1:2 where the “Spirit” (the same word as “wind” in Hebrew) hovers or blows over the waters. In Genesis 8, the text recounts how the dry land appears like was said in Genesis 1:9. Later on in Genesis 8:17, the text says, “be fruitful and multiply,” which is found in Genesis 1:28. Repeatedly, the Flood narrative links back with creation. It evidences that God is following through on His agenda established at creation. The Flood, while judgment, is part of God’s plan to restore things back to the way they ought to be. For this reason, God establishes the Noahic covenant (Gen 9:1–7). The term Noah means “rest,” and such rest is associated with Edenic rest, for the first time the root word occurs is in Genesis 2:15 which discusses that very reality [Mat96, pp. 208, 316–18]. With that, God’s agenda about creation is not merely theoretical; rather, while He judges the entire world, He also positions it toward Edenic rest. He will make all things right.
This agenda continues beyond the book of Genesis into Exodus. Within that book, the plagues themselves illustrate God’s control over creation. This is not merely conceptual. There are some deliberate links between the two. In the ten plagues, there are three sets of three plagues dealing with the water (Exod 7:14), the land (Exod 8:20; 9:1), and the sky (Exod 9:13; 10:1, 21). Those are the very categories that occur in Genesis 1 in the first three days of creation. In addition, the ten plagues use distinct language, such as “swarming,” found in Genesis 1 (Exod 8:3; cf. Gen 1:20). These allusions indicate that God’s work of creation and the plagues are tied together. It is the Lord’s declaration to Egypt and the entire world that He, and not the gods of Egypt (cf. Exod 12:12), is Creator and that He has an agenda to make all things right.

Such a context accounts for the details of another item in the book of Exodus: the Tabernacle. The very construction of the tabernacle resembles creation and Eden with, for instance, the blue tent representing the sky (Exod 26:1) [Bea04]. However, even the way the instructions for the Tabernacle is organized connects back with creation. The instructions repeat the language of “God said to Moses” (ֹדנֵּבֵּר יִתְהַ נִירָפַת) seven times and on the seventh time, the discussion is about the Sabbath (Exod 25:1; 30:11, 17, 22; 31:1; cf. 30:34; 31:12). The resemblance to the seven days of creation week where the seventh day is the Sabbath is striking [Kli77, Sai99]. The description of the tabernacle both in the building itself and even its literary presentation points the reader back to creation. God announces in the Tabernacle that while one is separated from God because of sin, there will be a day when man will be reunited with Him like in Eden. This not only reiterates the creational agenda but demonstrates that it does not apply to creation generically, but rather in details including worship.

Along that line, another item in Exodus goes back to creation: the Ten Commandments. There are numerous connections between creation and these imperatives. The first two commandments certainly uphold what God demonstrated at creation: He is the only true God and Creator, everything else is creation, and thus one should not confuse the creation with the Creator. The tenth commandment about coveting borrows language from Genesis 3 where the woman desired the fruit (cf. Gen 3:6).14 And the fourth commandment concerning the Sabbath directly connects back to God’s work at creation. The connection of the Ten Commandments back to creation reflect that the holiness God instituted at original creation is still the standard and Israel is to announce this to the world as a kingdom of priests (Exod 19:6). The world will not merely be made right physically, but rather its moral beauty and holiness will also be restored. God truly will make all things right.

Such an emphasis of holiness moves to the book of Leviticus. Within this, God states that when His holiness is fulfilled and Israel finally becomes a holy people, they will be His people and He will be their God. Within this, Leviticus 26:12 declares that God will walk amongst

---

14 See the term וַיִּבָּא הַמַּכְּר and compare with the phrase מַכָּר אֱלֹהִים found in Deuteronomy 5:21. The root מַכָּר is found in both.
His people. The language distinctively goes back to when God walks amongst Adam and his
wife in the garden. Again, God will restore the world back to Eden. And such a restoration
will be comprehensive, for it will even redeem man’s relationship with God. They will never
be lonely again but commune with God like Adam and his wife did in the garden.

Later books reveal the extensive nature of God’s sovereignty over all things as Creator and
thereby His determination to make all these things right. First Kings 4 is a good example
of this. That account lists the food the household of Solomon consumes in a day. In ex-
amining this list, it is interesting that only one other place in the entire Bible lists the same
animals: Deuteronomy 14. That earlier passage presents the clean foods which Israel may eat
when they enter the land. Initially, this confirms that Solomon’s kingdom is an indication of
how Israel was enjoying the promises that God has for them. However, it goes beyond this.
Deuteronomy 14 has the language of “you may eat any…” (cf. Deut 14:11, 20), which is very
similar language to what God said in the garden to the man (cf. Gen 2:16) [Tig96]. This im-
plies that Deuteronomy itself has ties back to creation and because 1 Kings 4 is connected with
Deuteronomy, it too has these ties. The bounty of Solomon’s kingdom reflects not only the
working out of God’s promises for Israel but the working out of God’s plan to restore what
was available in the garden of Eden [Demo3]. One day, there will be a time when creation is so
restored that even its food supply will be plentiful and there will be no hunger. The prophets
themselves tell of this (cf. Joel 3:18; Amos 9:14–15).

Speaking of which, the prophets themselves expound upon God’s creational agenda. Isaiah
11 says that the lamb will lay down with the wolf, and the bear will eat grass. Genesis
establishes that God directly created the animals (Gen 1:20–25), He is thereby sovereignly over
them, and therefore He will even make that aspect of creation right in the end. It goes beyond
this. In Psalm 51, David prays for God to create in him a clean heart (Ps 51:10). With that,
God’s creational agenda does not merely deal with things that are external and physical but
even that which is spiritual and internal. This too will be rectified because God is over all His
creation, both physical and spiritual.

While much more could be said of the Old Testament, a focal point of God’s creational
agenda occurs in the book of Daniel. In the center of the book, Daniel has a vision of the sky
and the sea, animals, and one like a son of man who rules over them all (cf. Dan 7:1–13).
Such a progression mirrors Genesis 1 [Laco1]. The vision demonstrates that what is true at
the beginning will be true at the end. This is the very substance of God’s creational agenda
discussed thus far. However, the one at the center of the vision is the one like a son of man,
the second person of the Trinity, the Lord Jesus Christ (cf. Dan 7:9–13).15 He is the final Adam
(cf. Rom 5:12) and will be the ultimate ruler of when God makes all things in His creation
right. Thus, the Old Testament shows not only the exhaustive ramifications of God’s work of

15 See discussion in [Cho13, pp. 131–39].
salvation but also its exclusivity. It all revolves around and centers upon His Son. All of this directly goes back to Genesis 1 and its design therein. God is Creator over all His creation and set one man over it all to rule it. Therefore, God, as Creator, will make all things right in His creation and set the God-man, His Son, to rule over it all in the end. The Old Testament has reiterated God’s agenda, covered it from a host of angles, and focused it upon Christ.

Such focus compels the storyline to move from Old Testament to New Testament. In the gospels, Jesus repeatedly calls Himself the Son of Man (Matt 8:20; 9:6; Mark 10:45; Luke 9:26; 19:10; Jhn 13:31) to accentuate that He is the final Adam prophesied in Daniel (cf. Luke 3:38). He has come to make all things right. And His miracles attest to this as He controls every element of this universe including plants (Jhn 2:1–12), the sea (Jhn 6:15–25), people (Matt 9:1–13), and even death (Mk 5:21–43). These signs, as the apostle John asserts, serve as a reminder that truly Jesus is the Creator, for in the beginning was the Word, and Jesus is that Word (Jhn 1:1). So Christ’s miracles demonstrate that Jesus is the Creator and how He has the sovereign might to restore things in His kingdom. Ultimately, our Lord’s death and resurrection are what secures this very restoration. He rises from the dead on the third day which is the first day of the week (Jhn 20:1). The emphasis on the first day of the week is significant, for it parallels the first week of creation. Christ’s resurrection begins a new era of creation. That is further supported by how the Lord is described post-resurrection. Mary mistakes Him to be a gardener (Jhn 20:15), an allusion to Adam the first gardener (Gen 2:8) [Kö9]. The timing and description of the resurrection are meant to parallel creation and show a new creation. Put simply, John’s gospel begins with creation in the beginning and it ends with a new creation. With that, the New Testament demonstrates that God’s creational agenda does not divert from what was established in the Old Testament, but rather climaxes, secures, and fulfills it.

The New Testament continues to show the ramifications of what it means that God will restore things to original creation. In Acts 2, at the birth of the church, there is language of creation associated with this new institution. The term “wind” found in Acts 2:2 (πνεῦμα) is a rare term in both the New Testament and the Greek translation of the Old. It is strictly associated with how God breathes in man the breath of life (Gen 2:7). With such language, Luke portrays the creation of the church like the creation of man, and the church then is a new humanity. This accords with the new creational work done by Christ, the final Adam, in the gospels, as well as with what Paul later says about the church as “one new man” (Eph 2:15) [Th10, p. 171], [Lin90]. This is also reflected in the apostle’s language of believers being a new creation (2 Cor 5:17) [Mar68, Har05]. Within this, in Acts 2, the church speaks in tongues. Given that the context is steeped in references to Genesis, speaking in tongues makes such a connection as well. It alludes to the first time the world ever spoke in tongues, which occurs in Genesis 11 at the Tower of Babel. There, man rebelled against God and He scattered them by dividing their language. Here, in Acts, man speaks in tongues because he has come back to God and God unites them. This new humanity is not merely transformed from the heart
but also united, something that has not happened since the Tower of Babel. The church is the proof that God will overcome sin and its effects, even the most pervasive ones. Accordingly, God will not only restore the physical world, animals, and one’s heart. He will even work such that there will be a new humanity united under a new Adam. Every detail and intent of original creation will be upheld. And that provides hope even for the question of whether there will ever be world peace. There will be because this is our Father’s world and He will make all of it right for His glory.

The church’s job is not merely to be transformed as part of God’s creational agenda but to point to the fulfillment of that agenda. Since the church is a first fruit of a new creation (Jas 1:17), they attest to a time for which creation groans (Rom 8:19–20); namely, when it will be set free from its slavery to corruption (Rom 8:21). So the church looks forward to where redemptive history is moving: when the Father will make all things in His world right. For this reason, as described in Revelation, God will eschatologically unleash a series of judgments. These acts of wrath have parallels to the creation week as the grass (Rev 8:7), water (Rev 8:9), and sky are affected (Rev 8:10) [Bea99]. God’s judgment is in a sense a de-creation to rebuild all things as He originally intended them to be. The complete proof of that occurs after Christ’s reign for a thousand years when Satan is released. Satan goes out to deceive the nations as he originally deceived Eve. Such a parallel shows that the point of this event is to answer the question: having restored the world, could the Fall ever happen again? However, unlike in Genesis 3 when Adam disobeyed and the world plunged into sin, the final Adam will never fail and so the redeemed world will never plunge into sin again. With that, God has made all things right and dealt with evil in such a way that accounts for everything—both individual issues as well as even the category of whether the Fall could ever happen again. In doing so, He has solved the problem of evil. And at that future moment, it is clear that this is the Father’s world, everything in it must be for His glory, and so He has made all things right in the end.

This discussion has commented on how God makes all things right, things that include the physical creation, morality, one’s relationship with God, food and hunger, loneliness, one’s spiritual heart, sin, the dysfunction of humanity, and evil itself. These are major questions people have concerning life and even ask about Christianity. Scripture does not shy away from these issues. Rather, the Bible has answers to these questions and the formulation of these answers revolves around creation. Because God has created this world for His glory, He will restore it all back to right. Redemptive history fleshes out the extent of this redemption, and it indeed is exhaustive. Put simply, every question people have about this world is answered by the doctrine of creation, for that doctrine declares that God is over this world. The doctrine of creation is foundational to answering the most pressing questions people have.

All of this, though, presumes that creation took place a certain way. It presumes God’s
direct act of creation and direct ownership of every part of it. It presumes that the created world existed in a certain state. It presumes that His creation was truly very good. If any of these presuppositions, established in Genesis itself, fall apart, then the entire logic of this full redemption falls apart. This highlights another danger of accommodation to evolution. It will undermine the very hope and resolution Christianity declares. However, more to the point, instead of being a liability to the veracity of Christianity, the doctrine of creation is actually one of its greatest assets. For in it is the grounds of why God does what He does and why we can have hope. We believe that all things can be dealt with because God had created a creation that was very good and not filled with death or decay. Since He is perfect and can make perfect things, He will do so in the end. Creation is not merely a doctrine of the past but a doctrine about the future. With that, as the queen of the sciences, theology and creation do not merely provide an authoritative framework for science, they also provide it with wonder, beauty, and hopefulness. For now, in this doctrine, science is the constant observation that this is our Father’s world, everything in it is for His glory, and He will make all things right in the end.

**Conclusion**

Modernist thought has been skeptical about the primacy of theology, and in this culture Christians have often thought of creation as a tertiary matter of little consequence. However, there is a reason that in the past theology was known as the queen of the sciences. Human reasoning, while impressive, can never trump divine revelation. Man is nothing in comparison with the infinite omniscient Creator of the universe. Job compellingly articulates this:

> The departed spirits tremble  
> Under the waters and their inhabitants.  
> Naked is Sheol before Him,  
> And Abaddon has no covering.  
> He stretches out the north over what is formless  
> And hangs the earth on nothing.  
> He wraps up the waters in His clouds,  
> And the cloud does not break out under them.  
> He obscures the face of His throne  
> And spreads His cloud over it.  
> He has marked a circle on the surface of the waters  
> At the boundary of light and darkness.  
> The pillars of heaven tremble  
> And are astonished at His rebuke.  
> He quieted the sea with His power,  
> And by His understanding He crushed Rahab.  
> By His breath the heavens are made beautiful;
His hand has pierced the fleeing serpent.
Behold, these are the fringes of His ways;
And how only with a whisper of a word do we hear of Him!
But His mighty thunder, who can understand? (Job 26:5–14)

God has created and runs the world that science discovers. Job declares that what is mysterious and puzzling to man, God established and knows. And what is astonishing is that the above list of God’s dominion over the natural and supernatural is just “the fringes of His ways…a whisper of a word….” What man endeavors to comprehend is the smallest sliver of God’s total knowledge. And all of this should humble man and reinforce that his knowledge is not as comprehensive or certain as he might think. Accordingly, what God has inerrantly revealed in Scripture is the complete, certain, and absolute truth. It stands far above what fallible, finite, and sinful man may derive. And this discussion has shown that, within this revelation, the doctrine of creation is not only clearly expressed but an inextricable part of the theology of Christianity. Far from being subordinate or irrelevant to man’s thinking, theology—specifically creation—is categorically above all human thought. It is the queen of the sciences.

It is fitting to discuss the issue of creation in the inaugural issue of a journal designed to explore the sciences from a Christian worldview. Starting points matter. And for science to be done right and to be used right requires that it starts right. And the doctrine of creation is that starting point. For this doctrine establishes why the sciences even exist and should have a crucial place in our study and society. The sciences engage in the discovery of all that is in our Father’s world. Their purpose is to illustrate how truly everything has been created for His glory. And though science will observe the fallenness of this world, because it continues to illustrate how this is our Father’s world, it points to the truth that God will make all things right. This is true science, one that is done according to truth, filled with worship, and offering true hope. Creation does not hinder science but recovers the nobility of science. It then cannot be an afterthought but the first thought for the sciences.

Bibliography


BIBLIOGRAPHY


32  THE JOURNAL OF THE MATH3MA INSTITUTE


A New Perspective of Entropy

Dr. Tai-Danae Bradley
PhD, Mathematics, CUNY Graduate Center
Research Mathematician at Sandbox@Alphabet
Visiting Research Professor of Mathematics at The Master’s University

Abstract: This article describes a new connection between two seemingly disparate topics in science, namely entropy and higher mathematics. It does not assume prior knowledge of either subject and begins with a brief introduction to information theory and a concept known as Shannon entropy, which we simply refer to as entropy. We then survey the vast landscape of higher mathematics, giving special attention to advanced analogues of high-school algebra and geometry known as abstract algebra and topology, respectively. Our goal is then to show that entropy, abstract algebra, and topology are inextricably linked through a version of a well-known formula from calculus known as the Leibniz rule. This result is given in the author’s recent work in [Bra21], and this present article is intended to give an overview of the ideas by gently introducing them from the ground up.

Introduction

In 2009 mathematician and theoretical physicist Freeman Dyson wrote an article for the American Mathematical Society in which he surveyed the works of notable mathematicians of the past few centuries. The scientific landscape is exceedingly broad, and yet as Dyson observed, mathematicians often fall into two categories [Dys09]:

Some mathematicians are birds, others are frogs. Birds fly high in the air and survey broad vistas of mathematics out to the far horizon. They delight in concepts that unify our thinking and bring together diverse problems from different parts of the landscape. Frogs live in the mud below and see only the flowers that grow nearby. They delight in the details of particular objects, and they solve problems one at a time.

Lest we conclude that birds are better than frogs or vice versa, Dyson quickly adds:
Mathematics is rich and beautiful because birds give it broad visions and frogs give it intricate
details.... It is stupid to claim that birds are better than frogs because they see farther, or that frogs
are better than birds because they see deeper. The world of mathematics is both broad and deep,
and we need birds and frogs working together to explore it.

The bird’s-eye view of the landscape is a valuable perspective, and discoveries of unexpected
connections between different parts of it are fascinating. But making those connections precise
and rigorous often requires a frog’s attention to detail. The subject of this present article has a
similar bird-and-frog feel to it. It is a new connection between information theory and parts of
higher mathematics related to algebra and geometry, and my recent technical article [Bra21]
contains all the “froggy” details. In this article, however, we will begin by flying high in the
air and surveying the ideas from a bird’s vantage point, occasionally landing on the ground
when necessary.

To begin, information theory fits broadly under the purview of science, technology, and
engineering, while more advanced versions of algebra and geometry (called abstract algebra and
topology, respectively) fit under higher mathematics. Information theory has a very applied
flavor, whereas higher mathematics has a very pure flavor. The two thus reside within separate
regions of the scientific landscape, and historically neither has had much much to say to the
other. But in recent years a few mathematicians have unearthed parts of what seems to be an
interesting bridge connecting them. Our present discussion is one of those small parts. It is
a new way to understand entropy from the perspective of higher mathematics. But what is
entropy? What does entropy have to do with information? What is the connection to higher
mathematics? And what is meant by “higher mathematics” anyway? We will answer these
questions one at a time.

A First Look at Information and Entropy

The study of information and communication finds its home in a branch of science called
information theory. At first glance it may be surprising to learn that information has its
own field of academic study, but a few moments of thought should dispel the surprise. After
all, what are the basic ingredients of communication? There must be an information source
(something that produces information), a channel (the medium through which information
is sent), and a destination (the person or object intended to receive the information) at least.
These simple ingredients quickly turn into a feast of questions. What if the channel has a
limited capacity? If some information is lost along the way, can it be recovered? If so, how
and to what extent? How is information stored, encoded, and decoded to produce meaningful
messages? Is it possible to quantify something as general as “information” in the first place?
Rephrasing these ideas in the precise language of mathematics allows such questions to be
asked and answered in more useful, quantitative ways. That is precisely what mathematician
and computer scientist Claude Shannon did in a seminal 1948 paper that launched the field of information theory [Sha48].

To see how mathematics can help quantify information, consider the following two statements: “The sun was shining in Los Angeles today,” and “There was a blizzard in Los Angeles today.” Which of those two sentences conveys more information? Readers familiar with US geography will know that it is almost always sunny in Southern California, so it is not surprising to learn that today was also sunny. Little information is conveyed in that first statement. On the other hand, it would be extraordinarily surprising—and somewhat distressing—to learn that Los Angeles was experiencing blizzard conditions. That would be a surprising scenario, and so a great deal of information is conveyed in the second statement.

These examples illustrate the intuitive idea that information and probability are inversely proportional. An event with high probability (“The sun was shining in Los Angeles today.”) seems to carry little information, whereas an event with low probability (“There was a blizzard in Los Angeles today.”) seems to carry lots of information. We can express this inverse relationship as a simple fraction. If an event occurs with probability $p$, then we might say the amount of information conveyed is $1/p$ because if $p$ is small, then $1/p$ is large and vice versa. This is almost the same quantity Shannon used in his 1948 paper, though he instead used the logarithm of $1/p$, which is a little more convenient to work with. This is a minor, technical detail for us, but let us briefly digress to explain what is meant by “convenient.” Think of an event that is 100% guaranteed to happen, that is, an event that occurs with probability $p = 1$. Since the event is certain to occur, it would not be surprising to learn that it did indeed happen. Intuitively, such a lack of surprise corresponds to the fact that no information has been conveyed. Zero surprise. Zero information. And yet since $p = 1$ the fraction $1/p = 1/1 = 1$ is not zero, which goes against that intuition. On the other hand, if $p = 1$, then $\log(1/p) = \log(1) = 0$ as desired. This is one reason why logarithms are more convenient. So, with this intuition in hand, we define the amount of information conveyed in a single event with probability $p$ to be the number $\log(1/p)$.

Thinking back to the weather, there are a range of possibilities—sunny, snowy, windy, cloudy, and so on. Each may occur in Los Angeles with a particular probability, so we may also compute the average (or “mean” or “expected”) value of information contained in a statement describing today’s weather. Generally speaking, this average amount of information has a name: entropy. More precisely it is called Shannon entropy to distinguish it from other notions of entropy that arise in science. (The precise formula for Shannon entropy will be given later on.) Perhaps the most familiar kind of entropy is that which appears in the Second Law of Thermodynamics, which says that the total entropy in a physical system never decreases. This kind of entropy is a measure of the amount of disorder or randomness in a system, and it

---

1 The logarithm of some number $x$ is another number that we will denote by $\log(x)$ (taken to be the natural logarithm in this paper). In particular, a useful fact to know in this article is that $\log(1) = 0$. 

is conceptually the same as Shannon’s version of entropy. Instead of asking about the weather in Los Angeles, we may instead ask about the speed or position of a molecule of gas, for instance. Other notions of entropy include von Neumann entropy, Tsallis entropy, Rényi entropy, and more. Our present discussion will not concern these, and so there will be no confusion if we simply refer to Shannon entropy as entropy. And notice that entropy, being the average of some numbers, is itself a number. It is not a vague notion or an intangible concept. It is a concrete mathematical object that has rich mathematical properties, as we will see in the page to come.

Entropy and information thus go hand-in-hand. Shannon’s entropy was introduced roughly 70 years ago in a quest for a mathematical theory of communication. Thermodynamic entropy has been studied since the 1870s. Both are still of great interest to scientists today. American theoretical physicist Lee Smolin once reflected on the role that entropy has played in the past and future directions of physics, from the discovery of atoms to modern-day research on black holes [Smo01]:

The search for the meaning of temperature and entropy of matter led to the discovery of atoms. The search for the meaning of the temperature and entropy of radiation led to the discovery of quanta. In just the same way, the search for the meaning of the temperature and entropy of a black hole is now leading to the discovery of the atomic structure of space time.

American-Israeli theoretical physicist Jacob Bekenstein, who died in 2015 and is known for his work on black hole thermodynamics, has also observed the fundamental relationship between information and the natural world [Bek03]:

Ask anybody what the physical world is made of, and you are likely to be told matter and energy. Yet if we have learned anything from engineering, biology and physics, information is just as crucial an ingredient.... Indeed, a current trend, initiated by John A. Wheeler of Princeton University, is to regard the physical world as made of information, with energy and matter as incidentals.

So entropy and information naturally arise in investigations of the physical world. One goal of this article is to show that entropy also naturally arises in higher mathematics—that is, in the sophisticated analogues of algebra and geometry alluded to above. It is natural, however, to wonder who might find these ideas interesting. Why is such a connection worth writing about?

Discovering the Unexpected

Nothing in this discussion so far suggests entropy may be related to the world of higher mathematics. Indeed, the only math used so far has involved fractions and probabilities. Higher, abstract mathematics is nowhere in sight. And yet, as we will see below, it is in fact inevitable. Entropy is therefore a link between two things that seem very different, which may suggest that deeper connections are waiting to be discovered. In a recent interview, Edward
Witten, one of the world’s premier mathematical physicists, reflected on his nearly 50 years of work in the field. When asked which current developments he is most excited about, his reply included entropy and related phenomena that may uncover some of the hidden mysteries surrounding the quantum world and Einstein’s theory of general relativity [Cha21]. It is intriguing to think about the mathematics that may underlie such developments. Speculations aside, this is simply meant to whet the reader’s appetite. Why would anyone devote time and attention to these technical ideas, that is, to a new way to understand entropy through abstract algebra and topology? Perhaps one day it may shed light on a new corner of science and mathematics not yet seen before. This is one reason why the ideas in the pages below are worth sharing. Consider what German theoretical physicist Max Plank, who won the 1918 Nobel prize in physics for his work in quantum theory, once said towards the end of his life [Pla48] (quoted in [Nic01, p. 201], emphasis added):

What has led me to science and made me since youth enthusiastic for it is not the at all obvious fact that the laws of our thoughts coincide with the regularity of the flow of impressions which we receive from the external world, [and] that it is therefore possible for man to reach conclusions through pure speculation about those regularities. Here it is of essential significance that the external world represents something independent of us, something absolute which we confront, and the search for the laws valid for this absolute appeared to me the most beautiful scientific task in life.

We will revisit this line of thought at the end of the article, but it is now time to turn to the mathematics. The next section will open with a brief introduction to the vast landscape of higher mathematics. Our attention will then shift to two regions within that landscape: abstract algebra and topology. We will give a short explanation of each, learning just enough to see how these kinds of advanced mathematical ideas are a natural part of entropy. The discussion will then climax into the main result of my technical article [Bra21], which describes a specific link between information theory and higher mathematics. We will then close with a final brief remark on the intrigue of such discoveries.

The Landscape of Higher Mathematics

The word “mathematics” may bring to mind the procedural material we once learned in school: word problems, long division, timed multiplication worksheets, and the like. But in reality the world of mathematics extends far beyond—and is very different from—the subject we are taught at a young age. So it is natural to wonder, “What does it mean to discover new mathematics?” Far from being a static subject, there is a sweeping, flourishing landscape of higher mathematics, and what is taught in school occupies only a tiny fraction of that land. It is elsewhere in this terrain that we will spend most of our time in this article. Afterwards, we will have seen an example of what it means to discover new mathematics. To that end, let us give a better description of our terrain.
The phrase “higher mathematics” becomes clearer when drawing an analogy with athletics. Upon learning that someone is an athlete, we may be curious to know which sport he or she plays. The word athlete is a broad term, and merely knowing that someone is an athlete does not tell us much about what that person does. The athletic landscape is comprised of a variety of sports, and any given athlete typically specializes in one or two of them: basketball, baseball, soccer, track and field, and so on. A professional athlete may go a step further and devote decades of his or her life to excelling in a specific role within a single sport. So, athletes generally differ greatly from one another even though they share a common profession. The same is true for mathematicians.

Like the athletic landscape, the mathematical landscape is also comprised of many different realms, and a professional mathematician may spend decades of his or her life in a particular locale within one of those realms. But unlike the names of our favorite sports, the names of these mathematical realms may sound less familiar: abstract algebra, topology, category theory, differential geometry, complex analysis, and more. And unlike the familiar features of the athletic landscape, the rolling hills and towering landmarks of the higher mathematical landscape go largely unnoticed despite their being foundational to science and technology. They are truly hidden in plain sight. There have, however, been some attempts at making this invisible landscape visible. One excellent example is the beautifully detailed map created by mathematician Martin Kuppe in the 2018 article [Bra18, p. 23]. The vintage-colored map entitled “Mathematician” takes the voyager on a quest through magical mathematical lands with cleverly devised names, such as “Probabilististan” and “Statistigrad,” the “Plains of Analysis,” and the “Ocean of Logic.” Two such regions within Kuppe’s map are most relevant to our present discussion on entropy, namely the “Califate of Al-Gebra” and the “Tundra of Topology.” The first is a play on words and refers to a branch of math called abstract algebra, which (as the name suggests) is a more sophisticated, abstract version of the subject we learn in high school. The second is topology, which is a more sophisticated, abstract version of geometry. Entropy is inherently algebraic and topological, so it will be helpful to first take a brief journal through both.

**Abstract Algebra: Math Beyond Numbers**

The word algebra originates from the Arabic word *al-jabr*, which means “reunion of broken parts.” It appears in the title of a ninth century book on the subject written by Persian scholar Mohammad ibn Mûsâ al-Khowârizmî [Gan26] and brings to mind the basic concept of combining things to form something new. If we have two numbers, for instance, then we can combine them, say by multiplication, to form a new number: $2 \times 3 = 6$. Of course there is nothing special about the numbers 2 and 3 in the previous sentence. If we have *any* two numbers, say $x$ and $y$, then we can multiply them to obtain a new number $x \times y$, also denoted by
This simple idea of combining things (whether by multiplication or addition or something else) quickly leads to the kind of algebra we learn in high school, where we are tasked with assignments such as “Simplify the expression \((x^3)^2y^4x^{-1}\) using laws of exponents” and “Factor the quadratic polynomial \(x^2 + 7x + 12\)” and the like. But this kind of algebra is nothing like the algebra studied at the graduate and research levels, which is called abstract algebra or modern algebra. To get a feel for the difference, it will help to think like a bird and not a frog. Forget the details. Forget the symbols. Forget words like “exponents” and “polynomial.” Instead, think back to the simple idea mentioned above: combining things to form something new. Whenever things can be combined, whether they are numbers or something else, there is likely algebraic structure behind the scenes.

Multiplication of numbers is just one example. Consider human language, for instance, where words combine to form longer expressions. Yellow is a word, and banana is a word, and we can “multiply” them to form the new expression yellow banana. The technical term for stacking words side-by-side is concatenation, as opposed to multiplication. But the term is not so important. The concept is. Concatenation and multiplication are conceptually similar. They both allow us to combine things to form something new. There is, however, a notable difference. The order in which we multiply numbers does not matter, whereas the order in which we concatenate words certainly matters. The product \(2 \times 3\) is the same as \(3 \times 2\), but yellow banana is not the same as banana yellow. This property is called commutativity. Multiplication of numbers is said to be commutative, whereas concatenation of English words is not commutative. Associativity, on the other hand, is a property shared by both multiplication and concatenation. When multiplying three numbers, it does not matter which two are multiplied first: \((2 \times 3) \times 5\) is the same as \(2 \times (3 \times 5)\), and the analogous holds for concatenation of English words.

This gives a taste of abstract algebra, where concrete details are abstracted away. In this branch of mathematics, the particulars of what is being combined, whether they are numbers, or words, or something else, is not the main focus. More important is the abstract structure, that is, the operation itself (multiplication or concatenation or...) and the properties it possesses. Informally speaking, any collection of things that can be combined—that is, where some notion of “multiplication” makes sense—is called an algebra, and when the multiplication possesses certain properties, the algebra is usually given a descriptive name. Examples include commutative algebras, associative algebras, Lie algebras, \(A_\infty\)-algebras, and more. In fact, an algebra is just one kind of algebraic structure. There are many more. Mathematicians also study groups, rings, fields, and vector spaces, to name a few. Each of these algebraic

---

\(^2\)More formally, an algebra is defined to be a vector space equipped with a way to multiply vectors. (These words will be familiar to students of linear algebra.) Lie algebras are named after Norwegian mathematician Marius Sophus Lie (1842–1899) and are used widely in physics. An \(A_\infty\)-algebra is one where the multiplication is not associative on the nose. Instead, it is only associative up to some wiggle room. This kind of structure appears in topology, the subject of the next section.
objects has a different (and sometimes multiple) notion(s) of combining things, and each plays a different role on the mathematical stage.

What, then, is the point of abstracting away details? Why pursue this line of thinking? One advantage is that it is clarifying. It helps us see relationships between things that initially seem unrelated. Karen H. Parshall, an American historian of mathematics, summarizes this nicely in an article on the history of abstract algebra in the *The Princeton Companion to Mathematics* [GBGL08, Section II.3]:

One objective of this new type of algebra is to understand the underlying structure of the objects and, in doing so, to build entire theories of groups or rings or fields. These abstract theories may then be applied in diverse settings where the basic axioms are satisfied but where it may not be at all apparent a priori that a group or ring or field may be lurking. This, in fact, is one of modern algebra’s great strengths: once we have proved a general fact about an algebraic structure, there is no need to prove that fact separately each time we come across an instance of that structure. This abstract approach allows us to recognize that contexts that may look quite different are in fact importantly similar.

With that, it appears we have progressed from high-school algebra to advanced mathematics rather quickly. Here is the bottom line. First, the claim that entropy can be understood in terms of algebra refers to *abstract* algebra and not to high-school algebra. Second, know that abstract algebra encompasses a large zoo of advanced mathematical structures. It is not necessary to know about any of them in detail, but it is good to be aware of their existence. This will help us make the connection to entropy later on. Here is a quick preview: it turns out that probabilities exhibit both algebraic and topological structure, and entropy interacts very nicely with both. More to the point, the *way* in which entropy interacts with algebra and topology is its defining characteristic—its fingerprint, so to speak. To understand this claim, however, we must first understand what topology is.

**Topology: Geometry’s Flexible Cousin**

Like geometry, topology is a branch of mathematics that involves the study of shapes. But unlike geometry, where angles, areas, lengths, and size take center stage, topology focuses on something else. What else can be said about shapes if not these features? *A lot.* Consider, for instance, the notion of sameness. What does it mean for two shapes to be equivalent? This may seem to be an innocent question, but as mathematician Barry Mazur once astutely observed [Maz07],

One can’t do mathematics for more than ten minutes without grappling, in some way or other, with the slippery notion of equality. Slippery, because the way in which objects are presented to us hardly ever, perhaps never, immediately tells us—without further commentary—when two of them are to be considered equal. We even see this, for example, if we try to define real numbers as decimals, and then have to mention aliases like $20 = 19.999\ldots$, a fact not unknown to the
merchants who price their items $19.99. The heart and soul of much mathematics consists of the fact that the “same” object can be presented to us in different ways.

A key difference between geometry and topology is how each answers the question, “When are two shapes considered the same?” In topology, shapes are thought of as malleable and pliable—made of something like Play-Doh—and two shapes are considered to be the same if one can be molded and deformed into the other without ever tearing or ripping the shape. As an example, the familiar shapes shown in Figure 1 are all considered to be equivalent. There is no difference between triangle, a square, a hexagon, or a circle in the eyes of topology. A triangle made of Play-Doh, for instance, can be deformed into a circle by smoothing the corners and rounding out the edges. So, rather than focusing on their differences—a triangle has three straight sides and a circle does not—topology instead embraces what they have in common: both shapes enclose a region on the page. On the other hand, a circle and a line are fundamentally different from this perspective. A line is not “closed” in the way that a circle is.

This notion of sameness is neither arbitrary nor meaningless mathematics, but rather is a consequence of formalizing another fundamental idea, namely that of closeness. In Mazur’s quote above, we understand that the number 19.99 is relatively close to 20, and that 19.999 is closer, and that 19.9999 is closer still. Said differently, the closeness of the numbers is measured by the distance between them. We also have the intuition that two items whose prices are close together will have values that are close together, as well. We could purchase a new book for $19.99, for instance, but would not expect that an extra penny could afford us a new luxury yacht for $20. The concept of mapping between objects (money and items in this case) in a way that preserves closeness is at the heart of topology. It is called continuity. More precisely, topology allows us to generalize the notion of distance in settings beyond numbers, and it does so in a way that also formalizes the idea of continuity. Simply put, any set—that is, any collection of things or elements—can be equipped with extra structure known as a topology. Very roughly speaking, “a topology on a set X” is a mathematician’s way of declaring which elements in X are close to each other. When considered together as a pair, both the set and its topology are referred to as a topological space. The formal definition is quite abstract and will not be given here, but it encompasses many familiar shapes. Lines, circles, triangles, squares, spheres, pyramids, cubes, and a plethora of more exotic shapes are all examples of topological spaces.

So, the field of topology generalizes distance or closeness, and this informs how one should
address the question of sameness. Two topological spaces are—again, roughly speaking—considered to be the “same” if they can be transformed into each other while preserving their respective versions of closeness. Such transformations include the deformations of Play-Doh triangles and circles mentioned above. While these ideas may sound quite different from high-school geometry, topology is a deeply rich and useful branch of mathematics, with applications ranging from explorations of the shape of space [Wee20] to DNA modeling [Adao04] to data analysis [CVJ21] and much more [Ghr14].

Entropy + Algebra + Topology = ?

Now that we have some familiarity with the higher mathematical landscape—algebra and topology in particular—we are ready to see how they are inextricably related to entropy. Our introductory discussion began with the observation that information and probability are inversely proportional. We also said that Shannon entropy, or entropy for short, is the average amount of information contained in a collection of probabilities. In other words, entropy is a number associated to a list of probabilities, and we interpret that number as a measure of information. That number is computed by a particular formula that will be shared below. The remainder of this article will rely heavily on that formula, and it will help to first establish some terminology in the next section. The subsequent pages aim to give a bird’s-eye view of the mathematics while occasionally providing frog-level details for interested readers. Such technical paragraphs are decorated with a triangle and labeled “▶ In more detail.” These paragraphs are included for the enjoyment of those who wish to dig deeper into the mathematics and may be safely skipped, if desired.

Entropy is a Number

We begin by introducing basic terminology. To start, a list of probabilities is called a probability distribution. That is, a probability distribution is a finite list of numbers between 0 and 1 whose sum is 1. For example, \( \left( \frac{1}{2}, \frac{1}{2} \right) \) and \( \left( \frac{2}{5}, \frac{1}{2}, \frac{1}{10} \right) \) are both probability distributions whereas \( (7, 3) \) and \( \left( -\frac{2}{5}, -\frac{1}{2}, -\frac{1}{10} \right) \) are not. In the first example, \( \frac{1}{2} \) and \( \frac{1}{2} \) may represent the respective probabilities of landing a heads or tails on a fair coin toss, while in the second example \( \frac{2}{5}, \frac{1}{2}, \) and \( \frac{1}{10} \) may represent the respective probabilities of choosing cereal, oatmeal, or fruit for breakfast. In both cases, it helps to think of probabilities as numbers associated to a finite set of options: the first option (choosing cereal), the second option (choosing oatmeal), the third option (choosing fruit), and so on.

Said more formally, given a natural number \( n \) (that is, a whole number \( 1, 2, 3, \ldots \)), a probability distribution on a set \( \{1, 2, \ldots, n\} \) is a list of nonnegative real numbers \( p = (p_1, p_2, \ldots, p_n) \)
satisfying \( p_1 + p_2 + \cdots + p_n = 1 \). Elements in the set \( \{1, 2, \ldots, n\} \) may be thought of as enumerating the different options or outcomes, each of which is assigned a particular probability. The letter \( p \) is being used as shorthand to represent the full list of numbers \((p_1, p_2, \ldots, p_n)\). Teasing this out with \( n = 2 \), suppose we have a set of two elements \( \{1, 2\} \) corresponding to the two faces of a coin, namely heads (option \#1) or tails (option \#2). If we let \( p_1 = \frac{1}{2} \) and \( p_2 = \frac{1}{2} \), then the list \( p = (p_1, p_2) \) is the first example of a probability distribution given above.

In addition to viewing probability distributions as lists, they can also be visualized with pictures. Figure 2 gives an example. There, the probability distribution \( p = \left( \frac{1}{2}, \frac{1}{2} \right) \) associated to a coin toss is represented by a stick-like tree with one root and two leaves, each representing a face of the coin. The words “tree” and “root” and “leaf” are technical terms used in a branch of mathematics called graph theory. (Even more formally, the pictures in Figure 2 are known as planar rooted trees.) The right-hand side of the figure shows a tidier version of this by labeling each leaf with its corresponding probability. We can likewise depict an arbitrary probability distribution \( p = (p_1, p_2, \ldots, p_n) \) as a tree with one root and \( n \) leaves that are labeled by the individual probabilities as in Figure 3. Visualizing probability distributions in this way will be a worthwhile adjustment for us, as we will see later on. A picture is worth a thousand words in mathematics, too.

What’s more, every probability distribution \( p \) has a number associated to it called entropy. This number, which we will denote by \( H(p) \), is given explicitly in the following formula.
Definition. The entropy of a probability distribution \( p = (p_1, p_2, \ldots, p_n) \) is defined to be

\[
H(p) = -p_1 \log(p_1) - p_2 \log(p_2) - \cdots - p_n \log(p_n).
\] (1)

It will be helpful to gain intuition for this expression, and we may start by comparing it to our opening remarks on entropy at the beginning of this article. There we defined the information contained in a single event with probability \( p \) to be the number \( \log(1/p) \). By basic properties of logarithms, this number is the same as \( \log(1) - \log(p) \) which is equal to \( -\log(p) \) because \( \log(1) = 0 \). In other words, the information in an event with probability \( p \) is the number \( -\log(p) \). But notice the formula for entropy in Equation (1) does not merely consider one event. It considers **multiple** events, each of which has its own probability. So the formula first computes the information associated to each event, that is, \( -\log(p_1) \) and \( -\log(p_2) \) and so on. Then it computes the average of those numbers by multiplying each by its respective probability and adding them. That is what is displayed in Equation (1), where the minus signs are important. The natural logarithm \( \log(p) \) is negative whenever its input \( p \) is between 0 and 1, so \( -\log(p) \) is nonnegative. In other words, the number \( H(p) \) is always either positive or zero.

Examples are helpful. Think of an event that is guaranteed to happen. An avid coffee drinker, for instance, will look forward to a cup of coffee each day. Suppose there is a 100\% chance they will have coffee each day and a 0\% chance they will not. This scenario corresponds to the probability distribution \( p = (1, 0) \). If we were to learn that this person indeed drank coffee today, then we would not be surprised. The behavior is expected, so no information has been conveyed. This lack of surprise is mathematically represented by the entropy of \( p \) in this example, which is zero:

\[
H(p) = -1 \log(1) - 0 \log(0) = -\log(1) = 0.
\]

Zero entropy corresponds to zero uncertainty. Now think about the other extreme, that is, an event with maximal uncertainty. Consider the outcome of tossing a coin. The result is either heads or tails, and neither is more likely assuming it is a fair coin. We expect the entropy of the corresponding probability distribution \( p = \left( \frac{1}{2}, \frac{1}{2} \right) \) to be positive since the outcome is totally uncertain. This is indeed the case:

\[
H(p) = -\frac{1}{2} \log \left( \frac{1}{2} \right) - \frac{1}{2} \log \left( \frac{1}{2} \right) = -\log \left( \frac{1}{2} \right) = -\log(1) - (-\log(2)) = \log(2).
\]

This computation is a special instance of a more general pattern. Whenever there are \( n \) outcomes each with equal probability \( \frac{1}{n} \), the entropy of the resulting probability distribution \( \left( \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n} \right) \) will always be \( \log(n) \), and this turns out to be the maximum possible value. In other words, one can show that \( 0 \leq H(p) \leq \log(n) \) for any probability distribution \( p \) on a set with \( n \) elements. So, entropy is akin to a measure of surprise or uncertainty, it has a formula,
and we have now seen two extremal examples. With these basics in hand, it is now time for a slight shift in perspective. We have seen that entropy is number, but in actuality, that number is a shadow of something more. Entropy is not merely a number. It is a function.

**Entropy is a Function**

Recall that every probability distribution \( p \) on \( n \) elements corresponds to a number \( H(p) \). The quantifier “every” suggests there is a function lurking behind the scenes. Indeed, the process of assigning a real number to a probability distribution is precisely what it means to have a function from the set of all probability distributions on \( n \) things to the set of real numbers. New mathematical notation allows us to restate this idea more conveniently:

\[
\text{For each natural number } n, \text{ there is a function } H : \Delta_n \to \mathbb{R}. \tag{2}
\]

We have used the letter \( H \) as the name of the function that assigns to a probability distribution \( p \) its entropy \( H(p) \). We will also use \( \Delta_n \) to denote the set of all possible probability distributions on the set \( \{1, 2, \ldots, n\} \). For example,

- \( \left( \frac{1}{2}, \frac{1}{2} \right) \) and \( (1, 0) \) are elements of the set \( \Delta_2 \),
- \( \left( \frac{4}{10}, \frac{5}{10}, \frac{1}{10} \right) \) and \( \left( \frac{1}{5}, \frac{3}{5}, \frac{1}{5} \right) \) are elements of the set \( \Delta_3 \),
- \( \left( \frac{2}{7}, \frac{1}{7}, 0, \frac{4}{7} \right) \) and \( \left( \frac{1}{7}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \) are elements of the set \( \Delta_4 \),

and so on. Mathematicians usually prefer to lower the index by 1 and write \( \Delta^{n-1} \) instead, but constantly “being off by 1” would be inconvenient in the work to come, so we will use \( \Delta_n \) instead.\(^3\) In our menagerie of mathematical notation, we have also used the symbol \( \mathbb{R} \) to denote the set of all real numbers. Moreover, the letter \( H \) paired with an arrow \( \to \) represents the assignment of a probability distribution \( p \) to the number \( H(p) \). The notation \( H : \Delta_n \to \mathbb{R} \) is very convenient, so we will always use analogous notation \( f : X \to Y \) to mean “a function \( f \) from a set \( X \) to a set \( Y \) that assigns to each input \( x \) in \( X \) one output \( f(x) \) in \( Y \).”

In summary, entropy defines a function \( H \), and we will refer to both \( H \) and its values \( H(p) \) by the same word: entropy. There is a subtlety, however. The claim that “entropy defines a function” is not the full truth. Entropy does not merely define one function. It defines infinitely many functions. The number \( n \) provides a clue.

---

\(^3\)The notation \( \Delta_n \) is pronounced “delta \( n \)” and is a clever choice. We will see later that the set of all probability distributions on three elements can be visualized as a triangle \( \Delta \).
Entropy is a Collection of Functions

Look back to the sentence displayed in (2): “For each natural number \( n \), there is a function \( H \).” The word order of that sentence implies that \( H \) depends on the value of \( n \). Since there are infinitely many natural numbers \( n = 1, 2, 3, \ldots \), there are necessarily infinitely many \( H \)s, as well. It would be helpful to indicate this dependency in our notation and write \( H_n \) instead of \( H \). Doing so would allow us to refine the sentence in (2) as follows:

\[
\text{Entropy defines a collection of functions } \{ H_n : \Delta_n \to \mathbb{R} \}.
\]  

(3)

This is better. Even so, the subscripts are rather cumbersome to carry around, so we will continue to omit them and write \( H \) instead of \( H_n \). Simply remember that \( H \) depends on \( n \). Notation aside, here is the essential idea: we must first choose a natural number \( n \) to specify the length of a probability distribution \( p \), and then we may compute its entropy \( H(p) \). In this way, entropy defines infinitely many functions, one for each natural number. There is \( H : \Delta_1 \to \mathbb{R} \) and \( H : \Delta_2 \to \mathbb{R} \) and \( H : \Delta_3 \to \mathbb{R} \), and so on.

Let us now pause and consider our changes in perspective. We began with the idea that entropy is a number. Then we observed that entropy defines a function. Now we see it defines a collection of infinitely many functions. There are many layers to entropy with still more to come. The functions \( H : \Delta_n \to \mathbb{R} \) turn out to possess nice mathematical properties, both individually and collectively. For instance, if we were to change the probabilities of a probability distribution \( p \) by a small amount, then it can be shown that the entropy \( H(p) \) would change by a small amount, as well. Conceptually this invokes the idea of “closeness.” Two probability distributions that are close or similar will have entropies that are likewise close or similar. This intuitive property has a name, which we briefly mentioned earlier—continuity. In other words, each of the functions \( H : \Delta_n \to \mathbb{R} \) are not merely functions; they are said to be continuous functions. We now find ourselves in the world of topology.

Entropy is a Collection of Continuous Functions

Recall that the symbol \( \Delta_n \) represents the set of all probability distributions on \( n \) things. Importantly, each of these sets \( \Delta_1, \Delta_2, \Delta_3, \ldots \) is not merely a set. There is a standard way in which they are in fact topological spaces. The first few are familiar and simple shapes. It can be shown that \( \Delta_1 \) is a point, \( \Delta_2 \) is a line segment, \( \Delta_3 \) is a triangle, and \( \Delta_4 \) is a pyramid, as displayed in Figure 4.

▶ In more detail. Consider the case when \( n = 1 \). The set \( \Delta_1 \) consists of all probability distributions on a single element. Such a probability distribution is a list consisting of a single number, and moreover that number must be equal to 1. So \( \Delta_1 \) is the singleton number 1,
which can be viewed as a point on the number line, as in Figure 4. A point is not a very interesting shape, but it is a shape nonetheless. Consider the more interesting case when \( n = 2 \). The set \( \Delta_2 \) of probability distributions on two elements is the set of all pairs of nonnegative numbers \( (x, y) \) whose sum is 1; that is, \( x + y = 1 \). We can rearrange this equation to see it is equivalent to \( y = 1 - x \), which is the equation of a line. Since we have further restricted \( x \) and \( y \) to be nonnegative, we may conclude that \( \Delta_2 \) is a line segment in the positive quadrant of the two-dimensional plane as in Figure 4. So, a point on that line segment is a probability distribution on two elements. Moving on to the case when \( n = 3 \), the set \( \Delta_3 \) of probability distributions on three elements is the set of all triples of nonnegative numbers \( (x, y, z) \) whose sum is 1; that is \( x + y + z = 1 \) or equivalently \( z = 1 - x - y \), which is the equation of a plane in three-dimensional space. Restricting to nonnegative coordinates gives rise to the triangular slice of the plane as in Figure 4. So \( \Delta_3 \) is likewise a very simple shape. It is a triangle, and a point on that triangle is a probability distribution on three elements. Understanding the case when \( n = 4 \) requires more work, but it can be shown that the set \( \Delta_4 \) of all probability distributions on four elements is a tetrahedron, or triangular pyramid, as shown in Figure 4. A point on that tetrahedron is a probability distribution on four elements.

The picture for \( \Delta_n \) becomes harder to visualize when \( n \) is greater than four, but each \( \Delta_n \) is indeed a topological space. The set of real numbers \( \mathbb{R} \) is also a topological space, and the upshot is that the functions \( H: \Delta_n \to \mathbb{R} \) defined by entropy are continuous with respect to the topologies. The \( \Delta_n \) further play an especially fundamental role in topology, where for each \( n \) the topological space \( \Delta_n \) is called an \( n-1 \)-simplex.\(^4\) For small values of \( n \) we have seen that simplices coincide with familiar objects: a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. In general, simplices are often used by topologists as building blocks for more complicated topological spaces, somewhat like Lego pieces. A hierarchy is already apparent: a tetrahedron is built up from triangles,

\(^4\)Recall that our \( \Delta_n \) are traditionally denoted by \( \Delta^{n-1} \), which is why we have called \( \Delta_n \) an \( n-1 \)-simplex rather than an \( n \)-simplex.
a triangle is built up from line segments, and a line segment consists of points. The popular online periodical *Quanta Magazine* recently expounded upon this idea [HE21]:

Many topological shapes can be built by gluing together pieces of different dimensions.... Individual pieces of the shape are grouped by dimension and then arranged hierarchically: The first level contains all the points, the next level contains all the lines, and so on. (There’s also an empty zeroth level, which simply serves as a foundation.) Each level is connected to the one below it by arrows, which indicate how they are glued together. For example, a solid triangle is linked to the three edges that form its boundary.

That “each level” is connected by arrows is an additional part of the mathematical theory that is not the focus of this article. But the quote above alludes to the fact that topological simplices are part of a larger framework that enables mathematicians to translate difficult topological problems into a language that is easier to work with. This framework is called homology, and the *Quanta* article reference above is a good place to learn more. It all starts by breaking up complicated topological spaces into little pieces or simplices. By definition, these same simplices have a probabilistic interpretation, and whenever there are probabilities, entropy is not far behind. Topology and entropy are thus inextricably linked. As we will soon see, algebra is just as inevitable.

Let us again pause to take inventory of our progress so far. We understand that entropy is not just a number but is rather a collection of infinitely many functions. Those functions moreover behave well from the perspective of topology because they are continuous. There now remains one final property of entropy to know about. One more layer to peel back. It is the fulcrum of this article and manifests itself whenever an event or outcome can be viewed as a composite process. This concept is best explained with an example.

**Composing Probabilities: Where Algebra and Topology Meet**

This section contains an example of “composing” probability distributions, an operation that will take center stage in our understanding of entropy. Both the example and ensuing discussion are heavily inspired by a 2011 informal article written by mathematical physicist John Baez [Bae11], as well as a talk given by mathematician Tom Leinster at the Centre International de Rencontres Mathématiques in 2017 [Lei]. The work and masterful expositions of both Baez and Leinster served as a primary source of motivation for the main result in [Bra21] that we are en route to unveiling, as well as the narrative we are sharing along the way.

Consider the following example. Suppose we flip a fair coin and then decide what to eat for breakfast or dinner depending on which face the coin lands. There is a 50-50 chance the coin will land on heads or tails, which corresponds to the probability distribution $p = \left( \frac{1}{2}, \frac{1}{2} \right)$. As shown previously in Figure 2, we may further represent $p$ as a (green) tree with two leaves labeled by the probabilities. Now, if the coin lands on heads, suppose we will choose what
to have for breakfast. Say there is a 40% chance we choose cereal, a 50% chance we choose oatmeal, and a 10% chance we choose fruit. This probability distribution on three breakfast items will be denoted by \( q = \left( \frac{2}{5}, \frac{1}{2}, \frac{1}{10} \right) \), which is a point in \( \Delta_3 \). The picture for this probability distribution is on the left-hand side of Figure 5. If the coin instead lands on tails, then suppose we will decide what to have for dinner. Say there is a 30% chance we choose pizza and a 70% chance we choose stir fry. Denote this probability distribution on two dinner options by \( r = \left( \frac{3}{10}, \frac{7}{10} \right) \), which is a point in \( \Delta_2 \). The picture of this probability distribution is shown on the right-hand side of Figure 5.

Notice there are five possible outcomes of this two-step process: cereal, oatmeal, fruit, pizza, or stir fry, depending on the coin toss. Importantly, each of those five outcomes has a probability associated to it, and those probabilities are easy to calculate. For example, the probability of flipping heads (a 50% chance) and then choosing cereal (a 40% chance) is the product of the probabilities of each individual outcome, namely \( \frac{1}{2} \times \frac{2}{5} = \frac{1}{5} \) or 20%. Similarly, the probability of flipping heads (a 50% chance) and then choosing oatmeal (a 50% chance) is equal to \( \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \) or 25%, and the probability of flipping heads and choosing fruit is equal to \( \frac{1}{2} \times \frac{1}{10} = \frac{1}{20} \) or 5%. This, too, has a visual counterpart. One can imagine grafting the root of the pink tree representing \( q \) onto the first leaf of \( p \), since the first leaf corresponds to flipping heads, and moreover letting the green probability \( \frac{1}{2} \) for “heads” propagate up through the three leaves of \( q \). Figure 6 shows the picture. Similar calculations show that the probability of flipping tails and choosing pizza is \( \frac{3}{20} \), and the probability of flipping tails and choosing stir fry is \( \frac{7}{20} \). The corresponding picture is analogous. To summarize this example, the process of flipping a coin and then choosing a meal gives rise to a new probability distribution on five things—cereal, oatmeal, fruit, pizza, stir fry—and based on our calculations above, that new probability distribution is equal to \( \left( \frac{1}{5}, \frac{1}{4}, \frac{1}{20}, \frac{3}{20}, \frac{7}{20} \right) \). It is a point in \( \Delta_5 \), and its picture is shown in Figure 7.

**Figure 5:** The probability distribution \( q \) on three elements can be represented as a tree with three leaves labeled by the probabilities, and similarly for \( r \).
on probability and statistics, for instance. Rather, it is an example of something new, and it is an essential ingredient in our discussion on entropy. To see how, it will be helpful to first introduce new notation for such composite probability distributions. Since this is a new mathematical construction, we are at liberty to invent our own notation for it. What should we choose? That is, what notation should we use to denote the probability distribution in Figure 7? We obtained the probabilities by multiplication, so it might be instructive to incorporate the “times symbol” \(\times\) somehow, which would also invoke the essence of abstract algebra introduced earlier, namely the notion of combining things to form something new. That is indeed what we have done here. A probability distribution \(p\) has been combined with probability distributions \(q\) and \(r\) to obtain a new probability distribution. We may therefore wish to denote this new distribution by, say, \(p \times (q,r)\) to remind us that the probabilities in \(q\) and \(r\) were multiplied by the probabilities in \(p\) to give rise to new probabilities. This would be a reasonable choice. Nevertheless, we will replace the \(\times\) with a circle \(\circ\) and write the following instead:

\[
p \circ (q,r) = \left( \frac{1}{5}, \frac{1}{4}, \frac{1}{20}, \frac{3}{20}, \frac{7}{20} \right).
\]

Why a circle? In a remarkable turn of events, this new mathematical operation was recently found to be not new at all. That is, our way of “multiplying” probability distributions—or rather, our way of composing them—is just one example of what can be represented by the tree grafting shown above. Many other kinds of composite mathematical objects fit neatly into the same template. That template is called an operad, and in the operadic literature a circle \(\circ\) is standard notation for the kind of composition seen in the example above. Moreover, our ability to compose probabilities is summarized in the fact that topological simplices \(\Delta_1, \Delta_2, \Delta_3, \ldots\) form...
So, what is an operad? The formal definition strikes a balance between the concrete and the abstract. It is concrete enough to be useful; it is abstract enough to subsume many examples. It is also rather technical and thus beyond the scope of this discussion. But some intuition will be helpful. Loosely speaking, an operad is an abstract mathematical tool that keeps track of certain properties of operations such as commutativity and associativity. One might think of these properties as “flavors” of multiplication. In the culinary world, there are many types of foods, and those foods come in a variety of flavors. Similarly, in the mathematical world, there are many types of operations and each may have a different flavor. Multiplication of numbers is just one example, but there are many more. Mathematicians frequently “multiply” things that are not numbers and then ask whether the flavors of those operations are interesting. We noted earlier that these more elaborate algebraic objects are well-known in mathematics with names such as commutative algebras, associative algebras, and Lie algebras, among many others. The language of operads distills these objects down to their core, leaving only the bare essentials that distinguish one algebra from another. For this to be useful, however, mathematicians had to first pin down an appropriate definition—one that was both rigorous and abstract, and this was accomplished in the early 1970s. An operad is defined to be a collection of “abstract operations” that accept $n$ inputs for each natural number $n = 1, 2, 3, \ldots$ together with a notion of composing them, and moreover that composition must satisfy a list of reasonable axioms. For an accessible introduction to the formal definition, see [Sta04, Bra17a, Bra17b] as well as [Lei21, Chapter 12.1].

**Figure 7**: There are five possible outcomes from flipping a coin and then choosing a meal. Each of those outcomes has a certain probability, which is obtained by multiplying individual probabilities from $p$ by those from $q$ or $r$. This composite process can be illustrated by grafting the respective trees as shown.
In more detail. Though a mouthful, the formal definition is conceptually simple. Pictures are especially helpful. An abstract operation generalizes the concept of a function that accepts \( n \) inputs and combines them to produce one output. That output can then be used as one of the inputs for a different function. Multiplication, for instance, is a function \( f: \mathbb{R}^2 \to \mathbb{R} \) that accepts a pair of numbers \((x, y)\) as input and computes their product \( f(x, y) = xy \) as output, which can be visualized by the cartoon in Figure 8. If we wish, we can then use the output \( xy \), which is just a number, as one of the inputs for some other function. This the basic idea behind function composition, a concept usually taught in high-school or college algebra. More generally, abstract operations can likewise be illustrated as cartoon-like trees with leaves that enumerate inputs and a single root to represent the output, exactly like those appearing in Figures 2, 3, 5, 6, and 7. In this way, we begin to see how probabilities start to fit into the theory of operads. But there is a mental exercise here. Probability distributions are now on par with abstract operations, which may seem confusing. How is a list of numbers an operation? It is not, really. It would be better to represent a probability distribution \( p = (p_1, p_2, \ldots, p_n) \) by an actual (continuous) function \( f: \mathbb{R}^n \to \mathbb{R} \) that accepts a list of numbers \( x = (x_1, x_2, \ldots, x_n) \) as input and that outputs a single number \( f(x) \). In such a passage from probability distributions to functions, the formula for \( f \) may involve the probabilities \( p_1, p_2, \ldots, p_n \) somehow. As an example, given a probability distribution \( p \) and a list of arbitrary numbers \( x = (x_1, x_2, \ldots, x_n) \) as input, we could define \( f(x) \) to be equal to the sum \( p_1x_1 + p_2x_2 + \cdots + p_nx_n \), a number known as the “dot product” between \( p \) and \( x \). In general, the passage from probability distributions to functions is a standard part of the theory of operads. Traditionally, such passages are (quite confusingly) called “algebras over the operad,” although one might prefer to call them representations of the operad [Bra21].

Needless to say, this topic is a specialized one. Not all mathematicians work with operads or are familiar with them, and yet the prevalence of operads throughout the higher mathematical landscape is quite astounding. Here is an overview given in [MSS02], a book written in 2002.
for graduate students, research mathematicians, and mathematical physicists:

Significant examples [of operads] first appeared in the 1960’s though the formal definition and appropriate generality waited for the 1970’s. These early occurrences were in algebraic topology in the study of (iterated) loop spaces and their chain algebras. In the 1990’s there was a renaissance and further development of the theory inspired by the discovery of new relationships with graph cohomology, representation theory, algebraic geometry, derived categories, Morse theory, symplectic and contact geometry, combinatorics, knot theory, moduli spaces, cohomology and, not least, theoretical physics, especially string field theory and deformation quantization.

While not all the terms may sound familiar, the variety is unmistakable. One more can now be added to the list: information theory. Around 2010, Leinster observed that the composition of probabilities described above is precisely what is needed to have an operad. That is, Leinster showed that the collection of topological simplices \( \Delta_1, \Delta_2, \Delta_3, \ldots \) admits the structure of an operad [Bae11, Lei21]. The upshot is that the way we have composed probability distributions \( p, q \) and \( r \) to obtain \( p \circ (q, r) \) in our coin-food example is neither homeless nor isolated in the land of mathematics. It finds a natural home in the established theory of operads. The composition of probabilities is moreover a collision between the worlds of algebra and topology. It is algebraic because we are combining probability distributions. It is topological because those probability distributions are elements of topological simplices. And as we will now see, entropy is not far behind.

The Chain Rule for Entropy

Recall that in its most basic sense, entropy is a number associated to a probability distribution. Our coin-food example involved four probability distributions—namely, \( p \) (a coin toss) and \( q \) (breakfast choices) and \( r \) (dinner choices) and their composition \( p \circ (q, r) \)—and each has an entropy associated to it as in Equation (1). Because the composite probability distribution \( p \circ (q, r) \) is built up from three individual distributions, it is natural to wonder whether the entropy of \( p \circ (q, r) \) can likewise be built up from the entropies of the three individual distributions. In other words, can the formula for \( H(p \circ (q, r)) \) be reexpressed in terms of \( H(p) \) and \( H(q) \) and \( H(r) \)? Perhaps, for instance, the entropy of the composite distribution is equal to the sum of the entropies of the individual distributions: \( H(p \circ (q, r)) = H(p) + H(q) + H(r) \).

While a good guess, this is not the case. But one can show the equality does hold if the entropies of \( q \) and \( r \) are multiplied by the probabilities of \( p \), as follows:

\[
H(p \circ (q, r)) = H(p) + \frac{1}{2}H(q) + \frac{1}{2}H(r). \tag{4}
\]

It may not be obvious why this modified equality is indeed true, but it can be easily verified using basic arithmetic and high school algebra. Simply apply the formula in Equation (1) to \( p \circ (q, r) \) and recall basic properties of the logarithm function.
Quite crucially, Equation (4) is just one example of a more general rule. An analogous equation holds whenever any probability distribution \( p = (p_1, p_2, \ldots, p_n) \) is combined with \( n \) other probability distributions, as in Figure 9, which will now be denoted with superscripts: \( q^1, q^2, \ldots, q^n \). (Pay careful attention to the difference between subscripts and superscripts: \( p_1 \) is a number, whereas \( q^1 \) is a list of numbers.) As in our motivating example above, the probability distributions that are composed with \( p \) may be of different lengths. For example, \( q^1 \) might be an element of \( \Delta_3 \), while \( q^2 \) might be an element of \( \Delta_{17} \), while \( q^3 \) might be an element of \( \Delta_5 \), and so on. In general, it can be shown that the entropy of the composite distribution \( p \circ (q^1, q^2, \ldots, q^n) \) satisfies the following important equation, which is sometimes called the **chain rule for entropy** [Lei21, Proposition 2.2.8]:

\[
H(p \circ (q^1, q^2, \ldots, q^n)) = H(p) + p_1 H(q^1) + p_2 H(q^2) + \cdots + p_n H(q^n).
\]

The chain rule is important because entropy is essentially the only collection of continuous functions that satisfies it. In other words, continuity and the chain rule are at the heart of entropy; they are enough to distinguish it from all other functions that assign real numbers to probability distributions. In mathematical parlance, entropy is said to be “uniquely characterized” by the chain rule. This is a theorem, and a proof of it may be found in a recent book by Leinster [Lei21, Theorem 2.5.1].

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.png}
\caption{A picture for the composite probability distribution \( p \circ (q^1, q^2, \ldots, q^n) \).}
\end{figure}

\begin{footnote}
\textbf{In more detail.} This characterization may be stated more formally as follows: if \( \{F: \Delta_n \to \mathbb{R}\} \) is any other collection of continuous functions satisfying Equation (5), then \( F \) must in fact be \textit{equal} to \( H \) or to some multiple of it. In symbols this means \( F = cH \) for some real number \( c \). As mentioned above, this important theorem appears in [Lei21] and is actually a slight variation of a closely related theorem about entropy that Russian mathematician Dmitry Faddeev proved more than six decades ago [Fad56]. Here is the precise statement of Leinster’s version of Faddeev’s theorem:
\end{footnote}
Theorem 1. Let \( \{ F: \Delta_n \to \mathbb{R} \}_{n \geq 1} \) be a sequence of functions. The following are equivalent:

i. The functions \( F \) are continuous and satisfy the chain rule.

ii. \( F = cH \) for some real number \( c \).

By now it might seem we are drowning in a sea of symbols. Perhaps our mathematical lungs are gasping for air. “Why does the chain rule hold? What is going on here?” These are reasonable questions. The first is easily resolved. Why is Equation (5) true? As claimed above, it can be verified using simple math. Walking through those calculations here, however, would cause us to sink deeper into the symbolic sea. But the details are found in [Lei21, Proposition 2.2.8] whose surrounding text also contains a more formal discussion of the mathematics in this section. Our second question is much more interesting. What is going on here? What does the chain rule in Equation (5) mean? To start, it tells us that entropy interacts with the operadic composition \( \circ \) of probability distributions in a very principled manner. Curiosity drives us to wonder whether a similar kind of interaction arises elsewhere in the mathematical landscape and, if so, where. On the surface this may appear to be a grandiose task. Where would we even begin to explore such a question?

A good place to begin is at the beginning. Suppose \( n = 1 \). Then the chain rule takes on the form \( H(p \circ q) = H(p) + H(q) \) for probability distributions \( p \) and \( q \). In words, this equation seems\(^5\) to say that entropy \( H \) behaves nicely with composition \( \circ \) of probability distributions and with addition \( + \) of real numbers. Functions with this kind of behavior may be familiar to those who have taken a course in abstract algebra, where such functions are called homomorphisms. To elaborate, in abstract algebra we often begin with a set \( X \) together with some way to combine or “multiply” elements in that set. This idea was introduced previously in this article, but let us unwind it further. Suppose \( \bullet \) and \( \bullet \) are any two elements of \( X \). These dots may represent numbers such as 2 and 3, or they may represent English words such as yellow and banana. We will not specify what the elements of \( X \) are because it does not matter. Supposing we can combine or multiply elements of \( X \), we will use juxtaposition \( \bullet \bullet \) to denote this operation. This could represent multiplication of real numbers, or concatenation of English words, or something entirely different. The abstraction allows us to represent all possibilities simultaneously. Similarly, suppose \( Y \) is another set with its own operation, which we will also denote by juxtaposition. Then a function \( f: X \to Y \) is called a homomorphism if combining elements in \( X \) and then applying \( f \) is the same as first applying \( f \) and then combining the elements in \( Y \); that is, if the function satisfies the equation \( f(\bullet \bullet) = f(\bullet) f(\bullet) \) for all elements \( \bullet \) and \( \bullet \) in \( X \). When comparing this with the chain rule for entropy when \( n = 1 \), it may seem that entropy is a homomorphism. The equation \( H(p \circ q) = H(p) + H(q) \) is indeed analogous to the newly

\(^5\) When \( n = 1 \) the only possible choice for \( p \) is the trivial probability distribution (1), which necessarily implies that \( p \circ q = q \) and \( H(p) = 0 \) and so the chain rule reduces to \( H(q) = H(q) \), which is uninteresting. But we will momentarily ignore this for the sake of exposition.
introduced equation \( f(\bullet\bullet) = f(\bullet)f(\bullet) \). So if we were asked to decipher the mathematical message in the chain rule, may we thus confidently assert, “Entropy is a homomorphism”?

Alas, the answer is no. Baez shared this conundrum in his 2011 article [Bae11]:

While [the chain rule] is cute, it’s a bit tricky to find its proper place in the world of abstract algebra.... Shannon entropy gives a map \( H \) from probability distributions to numbers. So, if you’re algebraically inclined, you would want \( H \) to be a homomorphism.... We see laws of this sort all over math. But the true law has an extra term. What’s going on?

To see where the problem lies, let us progress from \( n = 1 \) to \( n = 2 \). In that case, a probability distribution \( p = (p_1, p_2) \) may be combined with two other probability distributions \( q^1 \) and \( q^2 \), and the chain rule then becomes

\[
H(p \circ (q^1, q^2)) = H(p) + p_1 H(q^1) + p_2 H(q^2).
\] (6)

This makes it clear that entropy is not a homomorphism, as the \( p \)s have intermingled with the \( q \)s on the right-hand side. Alternatively, one might wish that the right-hand side of Equation (6) did not include the term \( H(p) \), for in its absence a small trick could be applied to make the equation look more like a homomorphism. (This is what Baez meant by “the true law has an extra term.” The details of the trick are given in his article.) So, it seems we are back to where we started. If entropy is not a homomorphism, then what is it? To gain the clarity we seek, we must strip away the details. It will help to squint our eyes while looking at the chain rule in Equation (6) and ignore most of the symbols. Forget the subscripts and superscripts. Forget the \( p \)s and \( q \)s. To see what is really going on with entropy, imagine that \( p \) is a green dot \( \bullet \) and the \( q \)s collectively are a pink dot \( \bullet \). Then Equation (6) roughly looks like something of the following form:

\[
H(\bullet\bullet) = H(\bullet) + \bullet H(\bullet).
\] (7)

In words, this says that the entropy of two things \( \bullet\bullet \) is equal to the entropy of the first thing \( \bullet \) plus the first thing \( \bullet \) multiplied by the entropy of the second thing \( \bullet \). Does this sound familiar? Perhaps not. Regardless, Equation (7) is undeniably asymmetric. It looks off, visually speaking. On the right-hand side of the equals sign there are two green dots but only one pink dot. That asymmetry is somewhat irksome, like having a pebble stuck in one’s shoe. The formula would appear more balanced if there were an extra pink dot on the right, like so:

\[
H(\bullet\bullet) = H(\bullet) \bullet + \bullet H(\bullet).
\] (8)

Now, does this look more familiar? Students of calculus may indeed recognize the equation above. It is reminiscent of a famous formula known as the “product rule” or the Leibniz rule, which is standard material in a first course on calculus.

\[\text{Notice this equation coincides with Equation } (4) \text{ when } p = \left(\frac{1}{2}, \frac{1}{2}\right) \text{ corresponds to a coin toss and when } q^1 = q \text{ and } q^2 = r \text{ correspond to our breakfast and dinner choices in the example in the previous section.}\]
In more detail. In calculus, the product rule is a formula for the derivative of a product of functions. To elaborate, the derivative of a differentiable function \( f : \mathbb{R} \to \mathbb{R} \) (that is, a function whose derivative exists) is another function often denoted by \( f' \) or by \( \frac{df}{dx} \) or by \( d(f) \). Given two differentiable functions \( f \) and \( g \), it is natural to ask about the derivative of their product. Can the derivative \((fg)'\) be expressed in terms of the individual functions \( f' \) and \( g' \)? The affirmative answer is given by the famous product rule, which says that the derivative \((fg)'\) is the function whose value at a point \( x \) is given by 
\[
(fg)'(x) = f'(x)g(x) + f(x)g'(x).
\]
In words, the derivative of \( fg \) is obtained by multiplying \( g \) by the derivative of \( f \), then multiplying \( f \) by the derivative of \( g \), and then taking the sum of these two functions. This may be written more succinctly as 
\[
d(fg) = d(f)g + fd(g),
\]
which should be compared with Equation (8).

But calculus is not the only place in mathematics where a version of the Leibniz rule appears. Many other functions may satisfy an analogous equation\(^7\) and such functions are given a name: derivations. Elaborating, suppose we have any set of elements \( \bullet, \bullet, \ldots \) together with some notion of “multiplication” between them so that we may make sense of expressions such as \( \bullet\bullet \). Perhaps these dots are numbers, but perhaps they are something else. The abstraction is once again intentional. Then, informally speaking, a derivation is defined to be any function \( d \) on this set that satisfies the Leibniz rule 
\[
d(\bullet\bullet) = d(\bullet)\bullet + \bullet d(\bullet)
\]
for all elements \( \bullet \) and \( \bullet \). As one might expect, this is a loose explanation of a much more formal definition, but the takeaway is that derivations are a staple in the world of advanced mathematics. A small digression may help to illuminate this claim. In our opening discussion on topological simplices, we briefly mentioned “homology,” which is a mathematical framework that assigns algebraic objects called homology groups to a topological space. Those groups encode valuable information about the topological space and are often easier to work with than the space itself. A similar story holds if the words “topological space” are replaced with “associative algebra.” There, the analogous construction is called the Hochschild cohomology of the algebra, and derivations of the algebra play a vital role: they are what are known as “cocycles of degree 1” [Wit19, Chapter 1]. Lingo aside, the takeaway is that derivations are functions that generalize the Leibniz rule from calculus, and Equations (7) and (8) hint at a tantalizing connection between derivations and entropy.

In more detail. As suggested above, derivations are functions defined with respect to some algebraic structure. That is, one must work in some kind of setting where “multiplication” makes sense. Given an algebra \( A \), for instance, a derivation on \( A \) is formally defined to be a

\(^7\) As an example, here is a simple exercise. Let \([0,1]\) denote the set of all numbers between and including 0 and 1, and define a function \( d : [0,1] \to \mathbb{R} \) by declaring \( d(x) = -x \log(x) \) if \( x > 0 \) and \( d(x) = 0 \) if \( x = 0 \). Show that \( d \) satisfies the Leibniz rule. That is, show that \( d(xy) = d(x)y + xd(y) \) for all numbers \( x \) and \( y \) in \([0,1]\). This computation only requires arithmetic and basic properties of the logarithm function.
linear function \( d : A \to A \) satisfying the Leibniz rule \( d(ab) = d(a)b + ad(b) \) for all elements \( a \) and \( b \) in \( A \). This can be generalized slightly by replacing the target \( A \) with another object \( M \) called a “bimodule over \( A \)” and instead considering functions \( d : A \to M \) satisfying the same equation. These, too, are called derivations. The only difference now is that \( d(a) \) is an element of \( M \), which is allowed to be different than \( A \). So, some care is needed here. If \( M \) and \( A \) are genuinely different from one another, then it is not at all obvious how to make sense of the expressions \( d(a)b \) and \( ad(b) \). Here, \( d(a) \) and \( b \) are elements of different sets, namely \( M \) and \( A \), respectively, and we have not said what it means to multiply elements that are not members of the same set. By way of analogy, multiplication of two numbers \( 0.5 \times 8 = 4 \) makes sense, but what would it mean to multiply a number with an English word, \( 0.5 \times \text{yellow} = ? \) This is the nature of the question we are faced with here, and a similar thought holds for \( a \) and \( d(b) \). Bimodules are the answer to such questions. A bimodule over \( A \) is a mathematical object \( M \) that is equipped with a way to “multiply” its elements by elements from \( A \). Readers familiar with linear algebra have seen this idea before. It generalizes what is known as scalar multiplication. The ability to multiply a vector \( v \) by a real number \( k \) to obtain a new vector \( kv \) is precisely the statement that a real vector space is a bimodule over the real numbers. In this analogy \( a \) and \( b \) are “scalars” and \( d(a) \) and \( d(b) \) are “vectors.”

In short, derivations are functions that interact with algebraic structure in a precise way known as the Leibniz rule. Moreover, the rough form of the chain rule in Equation (7) suggests that entropy behaves somewhat like a derivation. The similarity is indeed hard to miss:

\[
\begin{align*}
\text{Leibniz rule:} & \quad d(\bullet \bullet) = d(\bullet) \bullet + \bullet d(\bullet) \\
\text{entropy:} & \quad H(\bullet \bullet) = H(\bullet) + \bullet H(\bullet)
\end{align*}
\]

We are now in a position to ask the obvious question: Is there a real sense in which entropy is a derivation? Baez posed this very puzzle in his 2011 article, where he wondered how the similarity between entropy and derivations might be reconciled with some of Leinster’s work on the operad of topological simplices: “So an interesting question presents itself: How does the ‘derivation’ way of thinking about the [chain rule] relate to Tom Leinster’s interpretation of it...?” [Bae11] Ten years later, my work in [Bra21] gave an answer to this question.

The answer is that there is a correspondence—that is, a way to go back and forth—between Shannon entropy and derivations of the operad of topological simplices. The latter expression is a brand new generalization of the Leibniz rule in the context of operads, and a formal definition is one of the contributions of [Bra21]. The definition draws inspiration from the familiar concept of a derivation from abstract algebra, yet it is notably different. A derivation of an operad turns out to consist not of a single function, but rather of infinitely many functions. Happily, we have made a similar shift in perspective before. It is analogous to our understand-
ing that entropy is not merely a number, nor is it merely a function, but rather it is a collection of infinitely many continuous functions \( \{H : \Delta_n \to \mathbb{R}\} \) satisfying a certain equation—the chain rule. Analogously, a derivation of the operad of simplices is defined to be a collection of infinitely many continuous functions \( \{d : \Delta_n \to \mathbb{R}\} \) satisfying a certain equation—the Leibniz rule. The black box is a temporary place-holder for a new kind of output that we explain now. Recall that entropy assigns numbers to probability distributions. Our derivation, on the other hand, will assign functions to probability distributions. Explicitly, a derivation \( d \) of the operad of topological simplices assigns a continuous function \( d(p) : \mathbb{R}^n \to \mathbb{R} \) to each probability distribution \( p \) in \( \Delta_n \). We have seen an example of such an assignment already in a “representation of an operad” mentioned in our introduction to operads. Indeed, we noted previously that, because operads are abstract, it is better to work with concrete representations of them in practice. Our new version of a derivation takes this to heart. So to summarize, the first step in making the connection between entropy and derivations is to represent each probability distribution \( p \) in \( \Delta_n \) by a continuous function \( d(p) : \mathbb{R}^n \to \mathbb{R} \). And this gives a clue to the black box above. It is precisely the set of all continuous functions from \( \mathbb{R}^n \) to \( \mathbb{R} \). This set can further be made into a topological space, albeit one that is harder to visualize than simplices.\(^8\) 

Even so, let \( \text{hom}(\mathbb{R}^n, \mathbb{R}) \) denote this topological space of functions. Then we can summarize this discussion with the following informal definition.

**Definition (Informal).** A *derivation of the operad of topological simplices* is a collection of continuous functions \( \{d : \Delta_n \to \text{hom}(\mathbb{R}^n, \mathbb{R})\} \) that satisfies an appropriate version of the Leibniz rule, 
\[
\text{“}d(p \circ q) = d(p) \circ q + q \circ d(q)\text{”}
\]

for any probability distributions \( p \) and \( q \).

Of course, care must be taken to explain the desired Leibniz rule in the scare quotes above. The expression \( p \circ q \) does not make much sense, for instance. So far we have only used the symbol \( \circ \) to denote the composition of multiple probability distributions with an arbitrary \( p \) in \( \Delta_n \), as in the tree-grafting picture in Figure 9. To have an appropriate version of the Leibniz rule, however, we need only compose a single probability distribution with \( p \). But this problem is no problem at all. A single probability distribution \( q \) may indeed be composed with \( p \), and we have already done so in the example shown in Figure 6. That is, we can simply graft the root of some \( q \) onto any one of the leaves of some \( p \). The definition asks that an appropriate version of the Leibniz rule holds for each of those ways.

To make sense of the Leibniz rule in the context of the operad of topological simplices, we also need to make sense of the expressions \( d(p) \circ q \) and \( q \circ d(q) \) that appear in the formula. Notably, both expressions involve the combination of two things that are not the same; \( p \) is a probability distribution, whereas \( d(q) \) is a function. What does it mean to combine the two? We have been faced with this question once before. What kind of mathematical structure\(^8\)

\(^8\)Formally speaking, we can equip the set of continuous functions \( \mathbb{R}^n \to \mathbb{R} \) with what is known as the “product topology,” which turns it into a topological space that is easy to work with in this setting.
enables one to “multiply” elements from different sets in a meaningful way? As discussed above, the answer lies in a bimodule structure. Pinning down such details is another contribution of [Bra21]. The paper gives a formal definition of a “bimodule over an operad,” and it shows that the collection of topological spaces $\text{hom}([R^n, R])$ for each $n = 1, 2, 3, \ldots$ admits such a structure. Curiously enough, this part of the mathematics explains the reason that entropy appears to be missing a pink dot when compared with the traditional Leibniz rule in (9). See [Bra21, Example 3] for more details. So, a part of the mystery has now been solved.

With these definitions in hand, the rest of the mathematics falls into place as well. First, it can be shown that every derivation of the operad of topological simplices satisfies a version of the chain rule. This upgraded rule looks roughly like the following:

$$d(p \circ (q^1, q^2, \ldots, q^n)) \quad \text{sort of} \quad d(p) + p_1d(q^1) + p_2d(q^2) + \cdots + p_nd(q^n),$$

which is analogous to the original chain rule in Equation (5). The “sort of” hovering over the equals sign means interested readers are encouraged to take a look at the true equation, which is given in [Bra21, Proposition 1]. Either way, in our graphical notation this new version of the chain rule essentially says that the function $d(p \circ (q^1, q^2, \ldots, q^n))$ is obtained by applying $d$ to each of the trees representing the individual probability distributions. Below is a picture of this rule in the case when $n = 3$. The “dots” on the leaves can be ignored—they are part of the bimodule structure, whose explanation we omit.

Finally, with the proper definition of “derivation” in place, the main theorem of [Bra21]—and the climax of our discussion—follows immediately. One can show there is a way to go back and forth between Shannon entropy and derivations of the operad of topological simplices. More specifically, we can always use Shannon entropy to define a derivation and, conversely, every derivation knows about Shannon entropy. Here is the formal statement of the theorem:

**Theorem 2 (Bradley, 2021).** Shannon entropy defines a derivation of the operad of topological simplices, and for every derivation of this operad, there exists a point at which it is given by a constant multiple of Shannon entropy.

▶ In more detail. The statement of the theorem does not tell the reader exactly how the correspondence works, so let us provide a few more details for those who are curious. One direction is quite easy. To show that Shannon entropy defines a derivation, we need to use entropy to construct a collection of continuous functions $\{d: \Delta_n \to \text{hom}([R^n, R])\}$ and verify that it satisfies the appropriate version of the Leibniz rule hinted at above. Here is how to construct such a collection: for each natural number $n$ and for each probability distribution $p$ in $\Delta_n$, define

$$d(p \circ (q^1, q^2, \ldots, q^n))$$

where

$$d(p) + p_1d(q^1) + p_2d(q^2) + \cdots + p_nd(q^n),$$

which is analogous to the original chain rule in Equation (5). The “sort of” hovering over the equals sign means interested readers are encouraged to take a look at the true equation, which is given in [Bra21, Proposition 1]. Either way, in our graphical notation this new version of the chain rule essentially says that the function $d(p \circ (q^1, q^2, \ldots, q^n))$ is obtained by applying $d$ to each of the trees representing the individual probability distributions. Below is a picture of this rule in the case when $n = 3$. The “dots” on the leaves can be ignored—they are part of the bimodule structure, whose explanation we omit.

Finally, with the proper definition of “derivation” in place, the main theorem of [Bra21]—and the climax of our discussion—follows immediately. One can show there is a way to go back and forth between Shannon entropy and derivations of the operad of topological simplices. More specifically, we can always use Shannon entropy to define a derivation and, conversely, every derivation knows about Shannon entropy. Here is the formal statement of the theorem:

**Theorem 2 (Bradley, 2021).** Shannon entropy defines a derivation of the operad of topological simplices, and for every derivation of this operad, there exists a point at which it is given by a constant multiple of Shannon entropy.

▶ In more detail. The statement of the theorem does not tell the reader exactly how the correspondence works, so let us provide a few more details for those who are curious. One direction is quite easy. To show that Shannon entropy defines a derivation, we need to use entropy to construct a collection of continuous functions $\{d: \Delta_n \to \text{hom}([R^n, R])\}$ and verify that it satisfies the appropriate version of the Leibniz rule hinted at above. Here is how to construct such a collection: for each natural number $n$ and for each probability distribution $p$ in $\Delta_n$, define

$$d(p \circ (q^1, q^2, \ldots, q^n))$$

where

$$d(p) + p_1d(q^1) + p_2d(q^2) + \cdots + p_nd(q^n),$$

which is analogous to the original chain rule in Equation (5). The “sort of” hovering over the equals sign means interested readers are encouraged to take a look at the true equation, which is given in [Bra21, Proposition 1]. Either way, in our graphical notation this new version of the chain rule essentially says that the function $d(p \circ (q^1, q^2, \ldots, q^n))$ is obtained by applying $d$ to each of the trees representing the individual probability distributions. Below is a picture of this rule in the case when $n = 3$. The “dots” on the leaves can be ignored—they are part of the bimodule structure, whose explanation we omit.
the function $d(p) : \mathbb{R}^n \to \mathbb{R}$ to be constant at entropy; that is, define $d(p)(x) = H(p)$ for all points $x$ in $\mathbb{R}^n$. The proof that this defines a derivation is straightforward, requiring nothing more than arithmetic. The other direction of the correspondence is slightly more involved. It says that if $\{d: \Delta_n \to \text{hom}(\mathbb{R}^n, \mathbb{R})\}$ is any derivation, then for each natural number $n$ there exists a point $x$ in $\mathbb{R}^n$ so that $d(p)(x) = cH(p)$ for some real number $c$. The proof of this part of the theorem uses the result of Leinster-Faddeev mentioned previously in Theorem 1, which is the reason that both theorems involve a constant multiple of entropy. What’s more, one can easily show that the special point $x$ is actually zero, that is, $d(p)(0) = cH(p)$.

So, the mystery surrounding entropy and derivations is now solved. Or is it? We have reached the end of this article, but now there are new mysteries to explore. For instance, at the time of writing, it is not known whether there is a meaningful interpretation of derivations evaluated at other points besides zero, or whether the operad has other derivations besides the one that is constant at entropy. Further, it turns out that the original chain rule for entropy in Equation (5)—the very equation that caught the attention of Baez and Leinster [Bae11, Lei21]—is merely a corollary to the chain rule for derivations and the main theorem above. So, perhaps the chain rule is just a shadow of something more. Our “final layer” of entropy is almost certainly not the final layer at all.

**Venturing Into New Mathematics**

In closing, let us revisit a question asked towards the beginning of this article: “What does it mean to discover new mathematics?” We have now seen an in-depth example. We began with a survey of the landscape of higher mathematics and a basic introduction to Shannon entropy. We then began to peel back the layers of entropy one by one, culminating in the intriguing chain rule. Curiosity compelled us to ask where such a rule fit into the mathematical landscape, and this led us to pursue an interesting connection between entropy and derivations. Generally speaking, one way to approach such mysteries is to search the mathematical literature to see if the mystery has already been solved. If no such solution exists, then the mathematician is prompted to forge ahead. That was indeed the case for us. And that discovery—namely, a new way to think about entropy from a pure mathematical perspective—is the content of [Bra21].

But why is this new perspective worth sharing? Recall that information theory traditionally has little to do with either abstract algebra or topology, as each subject seemingly resides in separate, distant sectors of the scientific and mathematical landscapes. But now we have seen in great detail that entropy, algebra, and topology are intricately intertwined with one another. Recent events also indicate that the work in [Bra21] is merely one of several related connections between entropy and higher mathematics found in the past several years.
Perhaps these timely discoveries are a clue that there is more interesting, more fundamental mathematics waiting to be discovered. And because entropy is at the heart of it all, it is particularly intriguing to wonder about the new insights that such mathematics could lend to the study of physics and the natural world. In the words of German physicist Max Plank quoted earlier: “...the external world represents something independent of us, something absolute which we confront, and the search for the laws valid for this absolute appeared to me the most beautiful scientific task in life.” The beauty of such a task is especially enriched when the explorer has sure confidence that the answers may indeed be found.

Colossians 1:16–17

Bibliography


Imitation Learning: the Machine Learning Version of Discipleship

Dr. Monica B. Vroman
PhD, Rutgers University
Researcher at Rutgers University
Visiting Research Professor of Computer Science at The Master’s University

Abstract: Learning has been at the forefront of human activity for thousands of years. Machine learning, a much newer field, has recently gained popularity and gotten much attention in the media and in conversations around us. In this article, I define machine learning, describe its different branches, and explain how it works. I also answer some of the concerns and questions that people have about the future of technology, especially as it relates to advances in machine learning.

Introduction

Learning, the acquisition of knowledge and skill, has been at the forefront of human activity for thousands of years. In His common grace and kindness, God has created mankind in His own image and endowed him with a special ability to acquire and process information. This intelligence is one feature that sets men and women far above the rest of creation: consider the vast amount of learning required to even read and understand this article!

One outworking of image-bearing human intelligence and creativity is technology. An area of technology that has gained increasing popularity over the last decades is artificial intelligence (AI) and particularly the offshoot of machine learning (ML). Today, there is a plethora of adaptive robots and machines that change, learn, and adjust to their environments. God has made mankind in His own image and we are making machines in our own.

1 This introduction was coauthored with David Vroman, MDiv.
The advent of AI and machine learning has opened a new frontier of human knowledge and brought with it a multitude of important questions: In what sense can a machine learn? How does machine learning work? Should we be concerned or fearful about artificial intelligence and machine learning?

Unfortunately, in the minds of many, cinema has had more influence than science. Hollywood has painted vivid nightmares of cold and calculating tyrant robots ruling the world. According to these portrayals, humans will be surpassed by machines in every way, including intelligence, and we will be left at the mercy of merciless machines. But one need not be captivated by these doomsday scenarios to wonder about the dangers of artificial intelligence and where it might lead. In fact, the increasing prevalence of AI and machine learning coupled with rapid advances in the field do warrant careful thought.

For the Christian seeking to evaluate and submit every thought to the Truth (John 17:17, Proverbs 30:5, 2 Corinthians 10:5), thoughts about AI should be no exception. Our appraisal of AI and its potential must adhere to reality, and our responses to it ought to be shaped by the wisdom of the Bible. It is beyond the scope of this article to address spiritual responses to the realities and possibilities of AI or to address the entire breadth of AI. My goal in writing this article is to introduce the readers to the fascinating world of machine learning and to help them align their thinking with the present reality of ML. In this article, I will give the reader an idea of what machine learning is, how it works, and what are its current limitations.

This article will focus on a subset of ML called imitation learning. One of the most common and effective modes of learning is learning by example. Much of what is learned over the course of a lifetime is learned by observation and imitation. This reality can be seen in the practice of apprenticeship that has been common in many trades throughout history. It can also be seen in the pages of the Bible. The apostle Paul said, “Be imitators of me, just as I also am of Christ” (1 Corinthians 11:1, NASB). Jesus Himself regularly called on people to follow Him, to become His disciples, and to, themselves, make disciples (Matthew 28:19, 20).

In the rest of this article, we will define and examine how machines or computers are able to “gain knowledge” and in what sense they are able to learn.

**Machine Learning**

There is a common belief that machines, in general, are competent, exact, efficient, even infallible to a certain extent, and this view has been extended to machine learning as well. One day, my elementary school aged son proudly told me that he had successfully “fooled” his video game. This game uses the computer camera to assess how well children are able to copy drawings using game pieces of various shapes and sizes. He was missing a piece and was able to “trick” the program by making the required shape using other pieces than the ones prescribed. Another time, the same program did not accept a perfect copy of the drawing. Again,
the reaction was shock. Knowing how hard it is for computers to recognize objects from pictures, the subject of an area of research called computer vision, I had the opposite reaction. I was pleasantly surprised at how well the computer actually had done and that it only made those two mistakes. It struck me that my son and others see computer programs as an intellectual authority. But are they? Machines are very good at accomplishing certain tasks. Aside from their immense contribution to industry, we cannot imagine some of our daily tasks such as web search and using productivity software, without them. But they do have limitations. There are certain tasks, such as computer vision or recognizing various objects in a picture, that are extremely easy for humans and very hard for computers.

We will start by defining machine learning and what a machine is. A machine is “a mechanically, electrically, or electronically operated device for performing a task; [...] a computer” [MWb]. The “machine” in machine learning is a computer or a collection of computers. In other words, machine learning is a field concerned with how computers can be programmed to learn from the data they are given. For the remainder of this article, the term machine will refer to a computer or a computer program.

Since God has given humans the ability to learn so that they would know Him and enjoy Him forever (according to the Westminster Catechism [KRM86]) and since computers cannot know God, in what sense are computers able to learn and to what purpose? Going back to the definition of learning [MWa], we ought to ask, what are the processes involved in computers gaining knowledge? How do computers study, practice, and how are they taught? Can they have experiences? If yes, what are they and how do they learn from them?

Before we talk about computer learning, we need to understand how computers do what they do. Most of us use computers on a daily basis. They assist us with the work that we do, facilitate communication, help us stay organized, and help keep us entertained. They are useful to us.

Most people know that computers need to be programmed. Programming is a process that involves writing instructions using a syntax that depends on the programming language used. Like the variety of human languages, there is also a variety of programming languages, and they go by various names: Python, C++, Java, Julia, and so on. Writing code is similar to writing instructions in another language, except that language is very close to English and shares some of the same words with English. For example, if we want to program the computer to ask the user for two numbers and then display the number that is greater in the Python programming language, which I am choosing here because it uses a simpler syntax compared to other programming languages, we would write the code in Figure 1.

The words in dark red and italicised are words that the computer does not process and, therefore, we do not have to worry about whether the computer understands them or not. When we tell the computer print(text) the computer takes the text given between the parentheses and shows it on the screen. It does not peek at the text to see if it can understand
a = input("Please enter a number, a: 
")
b = input("Please enter another number, b: 
")

if b > a:
    print("b is greater than a")
else:
    print("a is greater than or equal to b")

Figure 1: Code for printing the larger of two numbers in Python.

it. The words in bold blue font in the above example are special words called instructions. These are words that the computer recognizes and knows how to obey or follow. Other words that the computer recognizes are called keywords and they are underlined in the code shown above. When the computer reads an instruction or a keyword, it has a predefined set of steps it executes to follow or accomplish that instruction. The human equivalent of a programming instruction is asking a human to do something. If we ask our roommate to “please get bread on the way home,” our roommate will follow certain steps to accomplish the task. Those steps might be to get in the car, drive to the store, park the car, get the wallet, get out of the car, lock the car, walk to the store, go to the bread aisle, take bread, pay for the bread, walk out of the store to the car, unlock the car, put the bread in the back seat, drive home. Thankfully, we do not have to describe all these steps to a human. We can just say something short such as “please get bread on your way home” and they bring bread when they come home. Similarly, it is helpful that we can tell the computer print("b is greater than a") and it shows the text "b is greater than a" on the screen. We do not have to tell the computer which of its switches need to be on or off or in which succession to do the steps it does to display something on its screen. Most of the computer instructions and keywords are in English and they are, typically, easy to recognize, at least to the average programmer. However, you may notice that the code I show above would be slightly awkward to read by a human and that is because it is not meant to be read by humans. Instead, it is meant to be a compromise between something that a human can communicate with the least amount of effort and something that a computer is able to understand. The awkwardness of programming languages comes from some of the fundamental differences between humans and machines, which the example of programming beautifully illustrates. On the other hand, machine learning bridges some of the gap between machines and humans. Understanding this gap, I believe, will help us appreciate some of what machine learning contributes to the field of computer science.

One of my favorite examples showing the difference in communication patterns between computers and humans is the peanut butter and jelly activity from CS unplugged [csu]. Like the other CS unplugged activities, the peanut butter and jelly activity does not involve the use of computers. Imagine I asked you to give me the instructions for making a peanut butter
and jelly sandwich and I faithfully wrote them down. Then, I got all the necessary ingredients and utensils and faithfully followed your instructions to a tee without any flexibility. Maybe you think that this activity would successfully result in a yummy peanut butter and jelly sandwich on the first try. You may be thinking, “What could possibly go wrong?” In practice, however, this example shows that, unlike programming languages, human language is not a strict and rigid tool. Rather, it is ambiguous, redundant, full of allegories, and analogies. Human language is dependent upon references to information that is common knowledge to humans such as “lemons are sour” or “animals do not drive cars.” Small children learn this “common knowledge,” but it would be extremely difficult if not impossible to encode in the “mind” of a computer [Dar20, IOM+21].

I highly recommend doing this activity with family or friends as a fun and entertaining game or looking up videos of people, typically teachers, who have done this activity in their classrooms. What this activity illustrates is that computers cannot handle ambiguity, figures of speech, exaggerations, sarcasm, etc. They take everything we tell them literally. Because of these features, programming languages are very formal, rigid, and every word has only one meaning. Programming a computer requires a certain level of skill that takes time and effort to learn. Websites such as CS Unplugged [csu], Hour of Code [hou], and others, make computer science ideas accessible to those who are interested in the big picture without the technical burden of learning programming.

While computer programming remains the main tool for getting computers to do what we want them to do, machine learning is takes a different approach: it does not involve feeding the computer the series of instructions that, when followed, lead to the computer accomplishing a certain task. Rather, machine learning algorithms describe the steps involved in how to learn to do the task from data. The qualitative difference between these two approaches is like the difference between following a recipe to make dinner, the equivalent of the classical programming paradigm, that is, following a set of instructions to accomplish a specific task, and learning how to drive by copying an expert driver. In the case of learning how to drive, we are learning the basics of how a car works and what its controls are, but then have the bulk of the learning come from just driving the car, seeing how the car works, and, at least in the beginning, getting feedback from the driving instructor. We could say that, in the driving example, the bulk of the learning uses the data we gather as we drive the car under various conditions and in various situations. By the time our drivers’ ed course is done, most of us are trained enough to know how to safely react to new situations based on the principles that we have learned under the supervision of a teacher. In a similar way, machine learning algorithms are programmed to learn from data or “experiences.” The task that they learn is not broken up into basic instructions that the computer can understand and know how to execute. Rather, the programming instructions make it so that the algorithm gets shaped by the data.
Definitions of Machine Learning. In 1959, well-known computer scientist Arthur Samuel, one of the first people who used the phrase “machine learning,” wrote a paper describing a machine learning algorithm that learned to play the checkers game based on the experiences of playing checkers against humans and against a copy of itself [Sam59]. His program was “told” what legal and illegal moves looked like and given a few basic rules needed to play correctly. Then, it learned winning strategies from the experience of playing the game against an opponent. Arthur Samuel defined machine learning as: “the field of study that gives computers the ability to learn without being explicitly programmed.”

In 1997, computer science professor Tom Mitchell wrote a book to introduce his graduate students to machine learning and give them the foundational knowledge they needed to do graduate-level research. In this book, he says, “Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience.” One way to think about this definition is to imagine a computer program that is learning how to play a game from the experience of playing against a copy of itself. You can think of this copy as the copy of the instructions that make up the program or a copy of the knowledge that the program has accumulated about the game since it started playing it. Examples of such information could be, what moves are best and what moves should be avoided in various situations or how many points the program receives when making various choices in the game. The performance measure tells us how well the program plays the game. One such measure is the percentage of games that the program has won. When we design a machine learning task, we need to define a few parameters, such as how will we measure the success of learning the task, how will we know that the task has been learned, what choices does the algorithm have, and what information will we give the machine learning program as it makes choices. In the checkers game, the available choices are the legal moves of the various pieces; the information available to the checkers player could be how many kings the player has and how many kings the opponent has or how many pieces it has left and how many pieces its opponent has left, or a combination of those and other information.

Just as a person gets better at a skill the more they practice that skill, so machine learning algorithms get better at what they do based on experience or on the data that is available to them. The idea of improvement with experience is at the core of machine learning and this same idea is what makes machine learning threatening because, if these algorithms can improve, how much can they improve? Is there a limit to their improvement or will they, one day, take over? To understand if machine learning is a threat to humanity, we need to understand a few basics about how these algorithms learn. Machine learning is a field that is as old as computers (for example, Arthur Samuel’s research on the game of checkers dates back to the late 1950s [Sam59]), but was never able to flourish to its full potential before this past decade because it was missing its main fuel, data. Machine learning is dependent on data. So, in the following, I will explain how computers learn from data.
Models. The concept of a model is central to machine learning. We can think of a model as a template, or a form that needs to be filled out. When we fill out a form, that form comes with some of its content already written. What we need to do is to fill in the blank information to complete the content of the form. Similarly, the structure of the model is determined ahead of time and the algorithms “fill in the blanks,” as it were, using the data they are given. Said more formally, a model is a function from a set of inputs to a set of outputs, and, moreover, that function is defined by certain numbers called parameters. Finding the right parameters is akin to filling out a form. I will give examples of models in the next few paragraphs.

There are four main branches of machine learning: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

Supervised Learning. Supervised learning methods are the most widely used. Some famous examples of supervised learning algorithms are neural networks, decision trees, support vector machines, and logistic regression. These algorithms learn from labeled data. Say we wanted to train an algorithm to recognize various fruits from pictures using supervised learning and that our data would be pictures of various fruits, each picture with a label telling the algorithm the name of the fruit (Figure 2). Supervised learning has two stages: training and prediction.

![Figure 2: An overview of the supervised learning process.](image)

During training, the algorithm learns the parameters of the model from labeled images of fruits. The goal of the training stage is that the algorithm would label a new image, which is the prediction or application stage in Figure 2.
As mentioned before, during training, a model is learned from the data. This process builds a mapping or relationship between images of fruits and words representing the names of those fruits (Figure 3). Most of the research in machine learning is focused on how this relationship is built. Imagine that the algorithm was trained using the data shown in Figure 3. One way to train the algorithm is to build a relationship between the colors in the images given in the training data and the labels of the images, a relationship that matches the training data. Imagine that the algorithm started by looking for the colors orange and green. If both were found, the algorithm would learn the label “mango” because the only fruit that is both orange and green in the training data is the mango. If the algorithm found orange and did not find green in the image, then the learned label would be “orange.” If the algorithm found red and green, the learned label would be “strawberry.” If red was found and not green, the learned label would be “apple.” It is easy to see that this method is not very useful because some pictures of oranges may also contain green if the picture shows a leaf; this algorithm would not be able to label kiwis since no fruit in the training data is both green and brown; apples are not just red, but they can also be other colors. But this method can be a starting point in understanding how the relationship between images and labels may be built. If training is successful, the learned relationship between images and words can accurately identify the name of a fruit in a new image (Figure 2).

Every supervised learning algorithm uses labeled data, but we are not always looking to predict a label, as in the example in Figure 2. In general, the data contains some information
that we are trying to predict. This information is called the dependent variables. In the case of labeling images of fruits, the dependent variable is the name of the fruit. The rest of the information in the data are the independent variables, which is the information given by the images in the example in Figure 2, for example, the pixels in the images. Labels are a special case of dependent variables.

The choice of the model or the template used usually depends on what we already know about the data. If we know that the relationship between the dependent and the independent variables is linear, that is, that when one grows, the other one grows at a similar rate, we could use a linear model. An example of data that could be modeled using a linear relationship is the connection between weight and height for a group of people. A linear model assumes that there is a known set of features for each training example (such as the height of a person) and that the relationship between these features and what we are trying to predict (the weight of a person) can be modeled by a line (Figure 4).

Figure 4: A plot of weight and height and the line that best models the relationship between the two [DoS21].

Limitations of Supervised Learning. Out of all the categories of machine learning, supervised learning has the greatest potential to advance the field and yield impressive results. Because of competition, companies using machine learning in their software do not make their methods public, so we do not know what data they have or what algorithms they use. Because of our limited knowledge of the best and most successful algorithms used today, it is difficult to gauge their limitations. What we do know is that the knowledge that these algorithms have is limited by the data used to train them. Google and Facebook only know about you what you "tell" them: they know what you search for, what you watch, the ads you see and those you
Unsupervised Learning. In unsupervised learning, we are not given the labels of the data. We may be given pictures of various fruits and told to extract patterns from those pictures. Imagine that you have never seen an orange, apple, or strawberry. If you saw pictures of these fruits, you would not be able to name them or describe their taste or smell. You may not even know their purpose and assume that they were not edible. This is what data looks like to unsupervised learning algorithms. There are no labels to learn. All that unsupervised learning can do is to discover patterns in the data. If we had never seen and tasted certain fruits, we would not be able to describe their taste or smell or know which ones make a tasty pie, but we would be able to sort them by color or shape. Some of the most popular unsupervised learning algorithms are called clustering algorithms and they group similar objects together, similarly to what I just described.

Limitations of Unsupervised Learning. Unsupervised algorithms are very limited because the data they receive is unlabeled, so it is very challenging to even gauge what has been learned and how useful it is.

Semi-supervised Learning. Semi-supervised learning is a combination of supervised and unsupervised learning. During its training stage, it uses a small amount of labeled data and a large amount of unlabeled data. Because it uses some labeled data, the performance of semi-supervised learning is, in general, better than that of unsupervised learning. Also, because of the small amount of labeled data it requires, semi-supervised approaches can learn from data that lacks labels by having a human label a small amount of the training data, a process that is less time-consuming than labeling the whole data set, which would be required in a supervised approach.

Reinforcement Learning. In reinforcement learning, the program that does the learning is called an agent. Unlike supervised learning, there are no inputs and outputs that the agent can use to learn the model. Instead, the agent receives a reward signal, that is, a positive reward for desirable actions and a negative reward (or cost) for undesirable ones. The agent is programmed to choose its actions in a way that makes its total reward as high as possible. Historically, reinforcement learning has been inspired by animal learning theories [AM06]. You can think

---

2 Sometimes things are a little more complicated than that. Some algorithms group users together based on various similarities and assume one piece of information on one member of a group is also true of the other members of the group. For example, recommendation systems, such as Netflix, group users together based on the kinds of movies they like. Say user A and user B are in the same group, they both like comedies and documentaries. Then, if A watches a movie that B has not watched and gives it a high rating, the system will recommend the movie to B.
of training a reinforcement learning agent in a similar way that a dolphin is trained, by giving it positive rewards every time it does something desirable.

The model used in reinforcement learning is called a Markov Decision Process (MDP) and it is used by the agent to learn how the world works. You can think of an MDP as a collection of objects. One of these objects is the collection of the states of the world, the equivalent of all the situations that the agent could find itself in. If the agent is learning how to park a car, the states may describe its distance from the various objects around it, its speed, etc. Another object in the MDP is the set of actions that are available to the agent. An agent who learns how to park a car would be able to choose to drive forward or in reverse, break, or turn the steering wheel. Another object in the MDP is a reward function, which tells the agent what is the reward signal it receives in each state, such as a high reward if the car is correctly parked and a negative reward or a cost if the car is too close to another car. Transition probabilities show the likelihood of the agent moving from one state to the next by taking actions. In a very windy environment of on an icy road, steering the car may take us in the direction we want to go with a high probability, or it may take us in a slightly different direction with a lower probability. The discount factor makes future rewards less appealing than present rewards, which “motivates” the agent to accomplish the task sooner rather than later. Formally, we would write that the MDP is a set:

\[ MDP : (S, A, T, R, \gamma) \]

where \( S \) is the set of states, \( A \) is the set of actions, \( T \) are the transition probabilities, \( R \) is the reward function, and \( \gamma \) is the discount factor.

The reinforcement learning agent uses an algorithm called value iteration to learn the value of each of its actions in every possible state. Imagine the small world shown in Figure 5. A smiley face shows the location where the agent is; in this case, the agent is in location 8. The states have been numbered and correspond to locations in this small “grid world.” The agent can move left, right, up, and down, and, every time it does an action, it receives a reward.
Imagine that the agent is in location 8 and takes the action to move to the left twice. These actions will take the agent to location 6, which is a puddle. In a puddle, suppose the agent receives a negative reward of $-10$ points. This reward signal is designed to discourage the agent from going into puddles. From location 6, suppose the agent takes the action to move to the left again. This action takes the agent to location 5 where, let's say, the agent receives a reward of 1000 points. This example illustrates how the agent is rewarded as it acts in the world. By convention in reinforcement learning, we assume that the agent does not know ahead of time how much reward it will receive in each location, so an initial random strategy or choice of actions makes sense. With every action, the agent learns how much reward it will receive by taking that action. After exploring the world by taking various actions in various locations, the agent learns what actions lead to ideal paths, that is paths that, when followed, yield the highest possible reward. For example, the value of location 8 can be calculated as the highest possible reward that the agent can get by following any path starting in location 8. After exploring the world enough, the agent learns that going through puddles has a cost and that the treasure chest corresponds to a high reward. The ideal path from location 8 is the one that maximizes the agent’s total reward. This path avoids puddles and ends up at location 5. An example of such a path is one that goes through locations 8, 7, 11, 15, 14, 13, 9, and 5. The value of location 8, therefore, is 1000, because that is the highest possible reward that the agent can receive by choosing its actions optimally from location 8.3

Limitations of Reinforcement Learning. Reinforcement learning algorithms have access to limited data. They get information about the environment they are learning in and are dependent on a reward signal. They are typically applied in specialized settings and the possibility of transferring their knowledge to new settings is very limited.4

In reinforcement learning, the agent is given the reward function and can learn an optimal strategy, that is, the choice of actions that maximizes its total reward. In imitation learning, the agent is not given the reward function. Instead, it estimates it by observing the behavior of another agent, considered an expert at the task to be learned and assumed to behave optimally in accomplishing that task. The learner’s reward function is a model that gets filled out as the learner observes the behavior of the expert. For this reason, imitation learning is also called inverse reinforcement learning. We will dig deeper into this process in the next section.

---

3 In this calculation, I used a discount factor of 1, which is the same as saying that a future reward of 10 is considered to have the same value as a present reward of 10.

4 To get an intuition on how reinforcement learning works and how long learning takes, watch this video on YouTube: “AI Learns to Park - Deep Reinforcement Learning” [Arz].
Imitation Learning: the Machine Learning Version of Discipleship

Let us consider an example that motivates my research in imitation learning. Suppose we want to design a robot that follows instructions given in plain English. Before the learning starts, assume the robot does not know the meaning of any English words. Instead, the system only knows the grammar (or the structure) of the instructions it will receive. For example, if we want to design a cleaning robot that receives instructions such as “Clean the garage,” the robot will know that the first word in the instruction is a verb or action and the rest of the sentence is the direct object, which is the object on which the action will be performed. It is important to note that the agent will start off not knowing what the words “clean” and “garage” mean, rather it will learn the meaning of words from examples of sentences paired with the behavior demonstrating what it looks like to follow those instructions. For example, we would tell the system: “Clean the garage” and demonstrate how to clean the garage. The system will watch us clean the garage, imitate our behavior, and learn what words in the sentence “Clean the garage” mean by assigning meaning to those words based on the task that was accomplished.

We show a simplified example in Figure 6. The agent demonstrating the tasks is shown as a gray diamond. Red objects are also textured with horizontal lines, yellow objects are textured with lines that go up and down, and cyan objects are textured with oblique lines.

In Figure 6, the first three columns are training instances. Each training instance is an instruction followed by an example of behavior. The behavior is shown as what the world looks like before and after the instruction is carried out or, what we call the beginning and end states. As mentioned before, the learning agent has been given enough information to know that “park” and “visit” are verbs or actions, “vermilion” and “sapphire” are adjectives, and “dish” and “plate” are nouns. The agent also has access to the world shown here as a grid world with colored and textured squares and other objects, where it can perform various actions such as moving up, down, right, and left, and learn their consequences, such as moving the agent.
shown as the gray diamond shape, up, down, right, and left. Some grid cells are colored and textured, which does not impact the movement of the agent, but some contain blocks, which are other objects with shapes, colors, and textures. If the agent attempts to move into a cell containing a block, that block moves in the same direction as the agent unless that block is impeded by another object or wall. In that case, no movement (of either agent or block) takes place. Using what it learns from the training instances, we would like our system to be able to receive the instruction in the fourth column, “Park the sapphire dish,” and carry it out using primitive actions in the domain from the given starting point.

Consider how one might solve this problem. Looking at the first training instance, we see that the agent moved to the red circle with horizontal lines and then pushed it over to the cyan cell with oblique lines next to the yellow pentagon with lines that go up and down. It is not clear from this one training instance what “park” means, but it very likely involves some combination of the red circle, the cyan cell, and/or the yellow pentagon, based on the agent’s proximity to these objects at the end of the action sequence. Looking at the third training instance, we see that the agent moved to the red pentagon and pushed it into the cyan cell. Generalizing across these two examples, it is reasonable to assume that “park” means “push to the cyan cell,” since this object appears in both end states near the end location, and the argument of “park” is a description of the object to be pushed. So, the “vermilion dish” is the red circle with horizontal lines in the first example and the “vermilion plate” is the red pentagon with horizontal lines in the third. Thus, “vermilion” refers to “red with horizontal lines,” “dish” to “circle,” and “plate” to “pentagon.” Next, we can look at the second example and see that the agent pushed the yellow circle with lines that go up and down out of the way and then proceeded to the yellow cell with lines that go up and down next to the blue pentagon. Because we have determined that “plate” refers to the pentagon, it would seem that the complement of “visit” describes an object to which the agent should become adjacent. Because the pentagon is blue, we can infer that “sapphire” means “blue.” Putting the pieces together, the command to “Park the sapphire dish” can be seen to mean “push the blue circle onto the cyan cell,” which can easily be carried out with a sequence of eight primitive actions.

Even though it may be easy or, at least, intuitive to conceptualize how humans such as small children could learn what language means from such examples, how would machines learn? If we use imitation learning as an example, the agent learner assumes that there is a reward function corresponding to each task that is illustrated by these examples. In other words, the behavior is generated by an agent teacher (the gray diamond in Figure 6) who is acting to maximize its reward according to its reward function. The learner does not know or have access to this reward function. Instead, it has access to demonstrations of the task to be learned. As mentioned before, this reward function is the model or template that the agent fills out as it observes the behavior of the expert modeling how the task is done. The template could be a certain type of function, for example, a linear combination of the features in the
grid world. These features could be the color of the squares that the agent traverses, how long
the path is that the agent takes, etc., and a linear combination of them is an expression of the
following form:

\[ R(s) = w_1 \cdot \text{cyan}(s) + w_2 \cdot \text{adjacentToBluePentagon}(s) + \ldots \]

where \( s \) is a state representing something akin to a location on the grid, and \( \text{cyan}(s) \) and
\( \text{adjacentToBluePentagon}(s) \) are numbers that can be interpreted as the amount of information
that is already in the form or template. For example, \( \text{cyan}(s) \) could be the number 1 if the
location \( s \) has the color cyan and 0 if the location \( s \) does not have the color cyan. The learner is
given this information already and does not have to learn it. What the agent needs to fill out
using the expert’s demonstrations are the “weights” of the features, or, the so-called reward
function parameters, \( w_1, w_2, \) etc. These are simply numbers that reflect how important each
feature is to the expert.

As the agent observes the expert, and as it assumes that the expert performs the task in an
optimal way, the agent learner can fill out this template by assigning a weight to each of these
features according to how important they seem to be to the expert. For example, if the expert
seems to be taking the shortest path to the goal, we can assume that each step costs the expert
something. Maybe the expert is factoring in the amount of battery it is using to accomplish the
task and trying to make it as small as possible, such that the expert prefers shortest paths to
the goal. If the expert seems to be going out of its way to step on red squares, we can assume
that it gets rewarded for each red square that it steps on, for example, if red squares are places
where it can charge its battery. If this is the behavior we are seeing, we can, in our template,
give a high weight to red squares based on the behavior of the expert.

**Maximum Likelihood Imitation Learning**

It turns out that the task of filling out the model for the reward function is not straightforward.
For any given behavior, one possible model is that the expert is always taking random actions.
This expert does not care about any of the information in its environment; it gets the same
reward no matter where it is and what it does. Even if it looks like the expert is trying to
take the shortest path to the goal, or go out of its way to step on red squares, since we do not
know anything about the expert’s reward function, it is always a possibility that the expert’s
strategy is to randomly choose its actions at each step and so, it happened by chance that this
expert made the choices we are observing. Because of this possibility, the imitation learning
task is said to be *ill-posed*. In other words, we do not have enough information to find the
expert’s reward function from demonstrations of corresponding optimal behavior. The way
to solve this problem is to add information that can help us arrive at the expert’s reward
function. This solution is accomplished by adding some assumptions to restrict the number of
possible reward functions that are explanations of the expert’s behavior. These assumptions vary among the different setups used in imitation learning research. For example, we could ask that the learner visits each state or location of the world about as often as the expert does. Or, we could ask that the policy or strategy of the learner be as close as possible to the policy or strategy of the expert. We can estimate the strategy of the expert by looking at the actions that the expert chooses in each state or location which we can estimate from their demonstrations. For example, we could look at how many times the expert visited a certain location during the demonstrations and count how many times the expert chose to go in each direction from that location. Then, our choice of action from that same location would follow the distribution of actions we have seen the expert choose.

Maximum likelihood inverse reinforcement learning (MLIRL), which was the focus of my graduate research [Vro14], estimates the expert’s reward function by adjusting the reward function parameters such that the likelihood of the expert’s demonstrations is as high as possible. As mentioned before, to solve the IRL problem, we need to make some extra assumptions. For example, we can assume that the expert’s reward function is a combination of known features. Take the grid world with puddles example. The features can be: ground, puddle, start, and goal (Figure 5). The expert’s reward function can be assumed to be a linear combination of these features, for example,

\[ R(s) = -0.1 \cdot \text{ground}(s) + (-10) \cdot \text{puddle}(s) + (-0.5) \cdot \text{start}(s) + 1000 \cdot \text{goal}(s) \]

where \( s \) is any state, and each feature \( \text{ground}(s) \), \( \text{puddle}(s) \), \( \text{start}(s) \), and \( \text{goal}(s) \) is true (1) if \( s \) is a location with just ground, a puddle, a starting location, and a goal location, respectively, or false (0) if \( s \) is not the location of ground, a puddle, the start, or a goal. For example, for location 5, the feature \( \text{ground}(s_5) \) is 1, \( \text{puddle}(s_5) \) is 0, \( \text{start}(s_5) \) is 0, and \( \text{goal}(s_5) \) is 1; the reward for \( s_5 \)

\[ R(s_5) = -0.1 \cdot \text{ground}(s_5) + (-10) \cdot \text{puddle}(s_5) + (-0.5) \cdot \text{start}(s_5) + 1000 \cdot \text{goal}(s_5) = 999.9. \]

If one was looking to interpret this reward function, it quantifies that the agent gets a high reward for reaching the goal. The agent also pays a small cost every time it takes a step toward that goal when it walks on the ground, and a higher cost when it walks in a puddle. To put it another way, the reward function parameters weigh each feature according to how much reward the agent gets in a state where the feature is true or how much it costs the agent to encounter that feature. Having a small cost for each step motivates the agent to take shorter paths to the goal.

In imitation learning we do not know the reward function, \( R \). Instead, we know the features of each state, whether it is a ground, a puddle, a start, or a goal. What we do not know are the parameters that weight each feature, the \( w_1, w_2, \) etc. We also know how the expert behaves and assume it is acting optimally with respect to its reward function. In other words, at each
step, we assume the expert chooses the action that maximizes its total reward. The goal of imitation learning is to find an estimate of the reward function that explains the behavior we are seeing from the expert. MLIRL finds the parameters of the reward function that maximize the likelihood of the observed behavior.

More specifically, MLIRL uses an algorithm called gradient ascent to find the reward function parameters (see also Figure 7). To understand gradient ascent, imagine trying to find the highest point in a given area, without being able to see around you farther than one step ahead. You start in an arbitrary location on a hill or on a mountain covered with fog. Your strategy is to look around you and choose to walk in the direction of the steepest incline. In other words, at each step, choose the direction that gets you as high as possible from this point in one step. This strategy is not guaranteed to take you to the peak of the mountain, but it is guaranteed to take you to a local peak, that is, a point that is higher than any of the points around it. The gradient ascent algorithm can get stuck in local maxima (the highest point in a certain neighborhood), instead of leading to the overall, general highest point. One solution to this problem is to repeat the algorithm and choose various starting points. This strategy could lead to the global peak.

Figure 7: Gradient ascent illustration from [Coa21].

The analogy of climbing a mountain works well in the case of two features and two parameters. For example, imagine a reward function $R(s) = w_1 \cdot \text{puddle}(s) + w_2 \cdot \text{ground}(s)$. For each pair of values for the two parameters ($w_1$ and $w_2$), we can determine the value of $R(s)$ since we know the features $\text{puddle}(s)$ and $\text{ground}(s)$ for each state or grid location, $s$. If we have $R(s)$, we can calculate an estimate of the expert’s policy and how likely the observed behavior is under that policy. These quantities are mathematical functions that depend on the values of the parameters $w_1$ and $w_2$. For example, for a given pair of values for $w_1$ and $w_2$, there is a function $L(w_1, w_2)$ that computes the likelihood of the observed expert behavior if $R(s) = w_1 \cdot \text{puddle}(s) + w_2 \cdot \text{ground}(s)$, since $R(s)$ is used to compute the policy, and the policy is used to compute the likelihood of the observed expert behavior. $L(w_1, w_2)$ is like the
altitude (or height) of the point given by latitude \( w_1 \) and longitude \( w_2 \). When we are at the point given by latitude \( w_1 \) and longitude \( w_2 \), we simply take the step of the steepest ascent and change the latitude \( w_1 \) and longitude \( w_1 \) accordingly. For reward functions with more than two features, the same principle and method apply, but it is harder to imagine the map in more than two dimensions.

This is how the MLIRL algorithm works:

1. Start at an arbitrary "location" (that is, choose arbitrary values for the reward function parameters, \( w_1, w_2, \ldots \)).
2. A set of values for \( w_1, w_2, \ldots \) gives us a reward function. This reward function is our current estimate of the expert’s reward function, \( R(s) \).
3. From the estimate of \( R(s) \), we can compute an estimate of the expert’s policy and the likelihood of the expert’s behavior under that policy, \( L(w_1, w_2, \ldots) \).
4. Take one step in the direction of the steepest ascent of \( L \). Change \( w_1, w_2, \ldots \) accordingly.
5. Repeat from step 2 until the parameters \( w_1, w_2, \ldots \) stop changing.

When the algorithm stops, we have reached a local peak. If desired, we can repeat the algorithm using a different starting “location” for the reward function parameters, that is, different starting values for \( w_1, w_2, \) etc. in Step 1, and see if we arrive at the same peak or at a different one. If we find various peaks, we choose the one that makes \( L \) as high as possible. When MLIRL is done, it found the parameters, and, therefore, the reward function, that maximizes the likelihood of the observed expert behavior.

Figure 8 shows, on the right, the reward function computed by MLIRL if the expert demonstrates the path shown on the left. This reward function assigns the highest reward to the location in the grid where the trajectory ends. The ground locations have a small positive reward of around 0.5 and the puddles have a negative reward of \( -1 \). It is interesting to note that the path shown is also the shortest path from S to G. It is possible that the expert did not care about puddles and it just so happened that, on the shortest path to the goal, it chose a path with no puddles. I have mentioned this ambiguity at the beginning of this section when
discussing that the imitation learning problem is “ill-posed.” MLIRL shows this ambiguity by choosing the reward function that makes the demonstrations most likely. In other words, the reward function shown on the right of Figure 8 makes the path shown on the left more likely than a reward function that would assign the same amount of reward to puddles and ground.

MLIRL Application in Carrying Out Instructions. Now, I will go back to the example I used to motivate imitation learning, that is, the example of learning how to carry out instructions in English from commands paired with demonstrations and, at the same time, learning the meaning of the words in the commands. As I mentioned when I introduced this problem, the commands are assumed to have a fixed form. In the example I showed in Figure 6, the grammar of the commands is: verb adjective noun. In other words, the first word is a verb in the imperative form, then comes the adjective describing the object, and, lastly, the name of the object. With this information, we could assume that each verb labels a certain reward function. In other words, there is a reward function associated with the verb “park,” another one associated with the word “visit,” etc. This link makes sense, since, as we have seen in the example, “park object” means “push object to the cyan square” and “visit object” means go to a square adjacent to the object. A next step could be to group together all the pairs of commands and demonstrations by the verb they use. We would put together all the pairs of commands and demonstrations that use “park,” all the ones that use “visit,” etc. Now, we have two demonstrations from the expert that is teaching us how to “park,” and one demonstration from the expert that is teaching us how to “visit.” We can use MLIRL to learn the reward function for “park,” and apply it to carry out the fourth instruction, “Park the sapphire dish.”

In my previous work [Vro14], I have used a simpler grammar to learn the meaning of words from demonstrations, a bag of words model which is also called a uni-gram language model. Language models are used to predict the probability of words in a sentence. In the uni-gram language model, every word has the same likelihood of appearing in the sentence and the order of words is ignored. With these assumptions, we can compute the probability of each word given each reward function. Let’s say that our training data are the ones shown in Figure 9. We have demonstrations of two tasks: moving the star object into the green room (Task 1), shown on the left side of Figure 9, and going to the green room (Task 2), shown on the right.

If we know that the instructions and demonstrations on the left side of Figure 9 correspond to one reward function and the instructions and demonstrations on the right side of Figure 9 correspond to a second reward function, we can compute the probability of seeing each word when the task is Task 1 and the probability of seeing each word when the task is Task 2, under the uni-gram language model. For example, we can compute the probability of seeing the
Figure 9: Training data for 2 tasks: Taking the star to the green room (left) and Going to the green room (right) [Vro14].

word "push" if the instructions correspond to Task 1:

\[ P("push"|\text{Task 1}) = \frac{2}{21} \] (11)

This number comes from the fact that we have three sentences instructing the agent to do Task 1, containing 21 words total. The word “push” shows up twice in these sentences, so the probability of seeing the word “push” when the demonstrated task is Task 1 is two out of 21, or 2/21.

From the demonstrations on the left in Figure 9, we can also learn the reward function for Task 1 and from the demonstrations on the right we can learn the reward function for Task 2, using MLIRL. Let’s say we receive the new instruction, “Go to the green room.” Which reward function should we use to follow this new command? Since we have the probabilities of each word given each task as shown in Equation (11), we can compute which task is this sentence more likely to fit. More specifically, we can compute the probabilities that this sentence would come from Task 1 and Task 2. The probability \( P("Go to the green room"|\text{Task 1}) \) that this sentence would come from Task 1 is equal to

\[ P("go"|\text{Task 1}) \cdot P("to"|\text{Task 1}) \cdot P("the"|\text{Task 1}) \cdot P("green"|\text{Task 1}) \cdot P("room"|\text{Task 1}), \]

and the probability \( P("Go to the green room"|\text{Task 2}) \) that this sentence would come from Task
2 is equal to

\[ P(\text{"go"}|\text{Task 2}) \cdot P(\text{"to"}|\text{Task 2}) \cdot P(\text{"the"}|\text{Task 2}) \cdot P(\text{"green"}|\text{Task 2}) \cdot P(\text{"room"}|\text{Task 2}). \]

We choose the task for which this probability is higher and apply its corresponding reward function to carry out the task.

You may see that this approach is very limited because the uni-gram language model is too simple to capture the complexities of human language, even of simple instructions such as the ones shown here. However, even with such a simple language model, MLIRL can be used to figure out reward functions from demonstrated tasks. With more complex language models such as the one using the simple grammar described in the explanation of Figure 6, I believe the system would be successful in interacting with humans and carry out commands on behalf of its users. Does this mean that I have successfully designed an artificial servant of humans who watches me do my work and learns to do it for me? Not quite. This is just the beginning of an exciting line of research, with many details that still need to be figured out.

Conclusions

In this article, I gave an overview of machine learning and what it means to learn in general and what it means for machines to learn. I defined what we mean by a “machine” that, in general, it is a computer programmed to learn from the data it is given. Next, I will summarize the answers to some of the questions I raised.

What are the processes involved in computers gaining knowledge? Machine learning algorithms learn from data. They have access to the knowledge embedded in the data they are given and nothing more. Depending on the structure of the data, the approaches used can be supervised, unsupervised, semi-supervised, or reinforcement learning. Each of these categories involves different processes, but they are all based on finding patterns in the data. Supervised approaches are most widely used and they involve learning a relationship between the data and what is to be predicted. Unsupervised algorithms extract patterns in the data. In reinforcement learning, the algorithm learns a strategy based on a model of the environment and a reward signal that rewards desirable actions and punishes undesirable actions.

How do computers study, practice, and how are they taught? From the data they are given, algorithms learn relationships, patterns, or strategies, depending on the problem they are solving. While computers do not literally study or practice, they do learn from data in that they can improve their performance by using the knowledge they are given. Supervised approaches learn relationships from labeled data, unsupervised algorithms are programmed to find patterns, and reinforcement learning agents are taught through a reward signal.
Can computers have experiences? What are they and how do they learn from them? Since data is the fuel of machine learning, the data given to these algorithms can be seen as experiences. Some of these data are given to the algorithm all at once, which is a typical setup for supervised and unsupervised methods. Reinforcement learning setups are, often, closer to what we, as humans, think of an experience-based learning: they take actions in their environment, receive rewards, and change their model of the world accordingly.

How much can machine learning algorithms improve? Is there a limit to their improvement? Machine learning is limited by the data that fuels it. It is difficult to answer these questions because most advanced machine learning algorithms today are not made public, including the data they use. So, it is difficult to know how much these algorithms can improve. But we can get an idea of how advanced they are by the applications that use them. The performance of self-driving cars, handwriting and voice recognition, face and object recognition, and so on, are all tasks that are handled extremely well by artificial agents today. One obvious current limitation of artificial intelligence and machine learning is that any one agent can make a lot of progress using the data it is given in one task or in a few closely related tasks. But, unlike humans, they cannot excel at multiple tasks that are unrelated. Another obvious limitation has to do with the fact that they do not have a soul or emotions. Their minds are simulated minds and they can only improve as far as our understanding of the human mind, which is very limited.

In my personal opinion, artificial intelligence will never surpass human intelligence. As I am looking at the perfections of our Creator, His perfect wisdom, intelligence, and power, I am struck by how we, humans, as wonderful, complex, and intelligent as He has created us to be, are so far below Him. He is infinitely above us in His excellence, holiness, wisdom, and judgement. On a much smaller scale, the machines we make are limited even more in their intelligence and in what they can do. Unless we perfectly understand how our minds and intellect work, we cannot even come close to building an intelligence that mimics ours.

Acknowledgements

I gratefully acknowledge Michael Littman and others for their involvement in conceptualizing the example illustrated in Figure 6.

Bibliography


Christ in Creation

Dr. John F. MacArthur
Chancellor of The Master’s University and Seminary
Pastor-Teacher of Grace Community Church

Featured Article: The following essay is adapted from John MacArthur’s talk at the first annual TheoTech Conference at The Master’s University on October 30, 2021 and is used by permission.

According to a recent article from Cornell University, astronomers have confidently affirmed that the universe came into existence 13.8 billion years ago—not 13.9, but 13.8 billion years ago—and subsequent to that came human life by chance [Gla21, C+20, Goh21]. The odds of human life appearing after the Big Bang have even been marked as one in a trillion trillion. Only after the earth came to exist did all the creatures appear. Finally, the humans evolved, but only after the exact sperm and egg cell evolved to generate a human with the DNA sequence that eventually brought us all into existence. Originally the chance of this was stated as one in 250 million for the sperm alone apart from the egg, and that needed to happen constantly and incessantly in an unbroken chain of millions of generations of your ancestors until it happened again by chance and created you.

Obviously, such a scenario stretches and strains credibility beyond the breaking point. It’s equal to stupidity. Evolution is built on myth, and the dominant myth of evolution is chance. Someone once said, “Chance is the new pillow for science to lie down on,” but the truth is that irrational explanation would keep a rational person awake all night.

The Myth of Chance

Reflecting on “the myth of chance” is helpful. Chance was once a word restricted to describing mathematical probability. It was later broadened to include any unpredictable event, any kind of probability or coincidence. But be clear: chance in itself is nothing. It is a word to describe
something else—other forces, other realities. Evolution has transformed chance into a force and a power. It has been elevated to be the greatest force in the universe; it is the power that created everything. Chance took nothing and turned it into everything. It took chaos and turned it into order. It took irrationality and turned it into coherence. It took “causelessness” and produced all effects. In fact, in an evolutionary world, chance is the creator. But this is, of course, logic and reason abandoned, which leaves only myth. The essence of myth cosmology is random, mindless nothing being the power that produces what is fixed law and rational reality. And the enemies of myth cosmology are empirical data and reason, which is the essence of pure science. If you neutralize either data or reason then you have enthroned myth, and that is exactly what scientists have done. They have created a hopeless schizophrenia.

Why does rational man do this? A twentieth-century Hungarian Jewish author named Arthur Koestler, who committed suicide with his third wife, offers this famous quote: “As long as chaos dominates the world, God is an anachronism” [Koe41]. And that’s the point. Getting rid of God leads to a kind of mental suicide that may lead to actual suicide. If chance is sovereign, God is not. And if God is not sovereign, God is not God by definition. But chance is not God, nor is it a force of any kind. Chance is nothing. It has no power to do anything. It can’t be the cause of anything. It can’t be the power behind anything. It can’t be the force that produces anything. Yet modern science, denying this truth, attributes incomprehensible power to chance. This hocus-pocus ignores law, science, logic, and reason, and it rejects the creative Latin phrase \textit{ex nihilo, nihilo fit}. Out of nothing, nothing comes.

When scientists attribute instrumental means to chance, they have affirmed absurdity. George Wald, a 1967 Nobel laureate, offered a telling quote from this perspective: “Given so much time, the ‘impossible’ becomes possible, the possible probable, and the probable virtually certain. One has only to wait: time itself performs the miracles” [Wal54]. So that’s how chance operates. Chance is a force that operates if there’s enough time. That in itself is another absurdity postulated by a man who was hailed as one of the great scientists in the world.

Quantum theory is a good place to see the folly of chance as a creative force. In 1900 German physicist Max Planck presented the idea that energy comes in discrete units called quanta. In 1927 Warner Heisenberg, another German physicist, found that when a photon strikes an atom, boosting an electron into a higher orbit, the electron moves from the lower to the upper orbit simultaneously without having traversed the intervening space. The electron simply ceases to exist at one location and simultaneously appears in another. Out of that came the “quantum leap.” It goes out of existence here, appears there, and never traverses the space in between. How is that possible? What is the power that makes that happen? There is no scientific explanation, hence the confusion of being able to explain it. In truth, the reality that explains this is found in what the Scripture says about Christ who “upholds all things by the word of His power” (Hebrews 1:3). There is a divine person who holds all atomic energy together. God is in the gaps. Chance and time can’t produce anything. (In fact, how can there
Early computer scientist John von Neumann once imagined what became known as the von Neumann machine. He said that the perfect machine would be self-generating, self-repairing, and self-reproducing, but no one could produce it because it would be too complicated. However, that’s exactly what every living cell is: self-generating, self-repairing, and self-reproducing, and God has made trillions of them in each human body. Also think back to the Middle Ages when theology was called “the queen of the sciences”; that is, theology ruled science. Theology provides the only rational basis for the existence of anything, and Genesis 1–3 is the rational account of creation by the Creator Himself. Everything had to start with the uncreated God. Without God, all explanations are varieties of the absurd. In Exodus 3 God says, “I AM THAT I AM,” indicating that there was never a time when God did not exist. He is self-existent, self-sufficient, and sovereign. And in Genesis where the I AM acts to create, there is a simple and rational declaration: “In the beginning, God created the heavens and the earth” (Genesis 1:1).

Anthropologist Herbert Spencer, who died in 1903, identified five categories of the knowable: time, force, action, space, and matter [Spe67]. Everything, he claimed, fits into one of those five categories. But long before Spencer, Genesis 1:1 made the declaration of those very categories: “In the beginning” (time), “God” (force), “created” (action), “the heavens” (space), “and the earth” (matter). So the first verse of the Bible covers all the categories. There is only one account of the creation, and that is the verbal testimony of the Creator—His spoken intelligent words. This is far from the ridiculous notion that nobody times nothing equals everything. It is far from the absurdity that chance, which is nothing, was the power of creation. And it is far from the irrational notion that time existed before there was anything. There’s only one reasonable reality, and that is the fact that the eternal God is the Creator.

In the Beginning Was the Word

The creation account in Genesis is further detailed in John 1:1–5 [LSB21]:

In the beginning was the Word, and the Word was with God, and the Word was God. He was in the beginning with God. All things came into being through Him, and apart from Him nothing came into being that has come into being. In Him was life, and the life was the Light of men. And the Light shines in the darkness, and the darkness did not overtake it.

This revelation introduces something highly significant. The Creator is identified as “the Word,” which appears three times in the first verse: the Word was God, was with God, and was in the beginning with God. Moreover, all things were created by the Word, in the Word was life, and “the life was the Light of men.” There is no explanation in this text as to why the phrase “the Word” is used the way it is, because it was already familiar to the readers of the New Testament—both to Jews and Greeks. Greeks were familiar with the term logos (word),
which some identified as the supernatural, impersonal power and mind behind creation. They recognized that information was essential to creation. And the Jews knew what the Old Testament said. Consider Psalm 68:32–33, for instance: “Sing to God, O kingdoms of the earth, sing praises to the Lord, to Him who rides upon the highest heavens... behold, He speaks forth with His voice, a mighty voice.” God created with His voice, with words. Psalm 29:3–11 further says [LSB21]:

The voice of Yahweh is upon the waters; the glory of God thunders, Yahweh is over many waters. The voice of Yahweh is powerful, the voice of Yahweh is full of splendor. The voice of Yahweh breaks the cedars; indeed, Yahweh breaks in pieces the cedars of Lebanon. He makes Lebanon skip like a calf, and Sirion like a young wild ox. The voice of Yahweh hews out flames of fire. The voice of Yahweh causes the wilderness to tremble; Yahweh causes the wilderness of Kadesh to tremble. The voice of Yahweh makes the deer to calve and strips the forests bare; and in His temple everything says, “Glory!” Yahweh sat enthroned over the flood; indeed, Yahweh sits as King forever. Yahweh will give strength to His people; Yahweh will bless His people with peace.

The eternally existing One spoke everything into existence by His words and sustains it all by His words. This is also reiterated in 2 Samuel 22:14: “The Lord thundered from heaven, and the Most High uttered His voice.” These passages and many more show that the word of the Lord is a significant Old Testament theme. The word of the Lord came again and again because God spoke. In other words, God is a speaking God, which means He is intelligent, and He is a communicator. The word of the Lord was from God; it was His communication coming with power to bring into existence whatever His will desired.

Thus, we know by this that the Creator is a communicator. The Creator possesses reason, cognition, massive intelligence, wisdom, and incomprehensible power. In ancient times, common people in the Greek world understood the logos as a force behind what existed, and the Jews knew it was the word of the Lord behind all that was created. The apostle John presents the truth that the creative power, the creative force, and the creative wisdom is not an impersonal reality, but is the personal God who came into the world in the man Jesus: “The Word became flesh, and dwelt among us, and we saw His glory, glory as of the only begotten from the Father, full of grace and truth” (John 1:14). So, God spoke and all that exists came into existence by His words.

This foundational point is even observable from the vantage point of science because what precedes the creation and function of anything is information. Wallace B. Henley has written an interesting book on artificial intelligence in which he asks the question posed in the book’s title: Who will rule the coming gods? The thesis of the book is that godless society will create machines more intelligent than a person. Those machines will take the place of God. So he asks, “Who will rule the coming gods of artificial intelligence?” The book also considers the role that information plays in bringing anything into existence. Here is an illustration [Hen21]:

...during the Second World War, great effort was made to break the Enigma code used by the Germans to pass secret information. A code-breaking project was established at Bletchley Park
in Buckinghamshire, England. The goal of Alan Turing and his team was to break the code the Germans changed daily through the Enigma device which could configure messages in 150 million million ways.

The Bletchley machines existed only because the information coded and transmitted through Enigma already existed. Thus, in the sequence of data processing, there was the primacy of information. The information called forth the existence of the machines. The devices and systems that Bletchley made didn’t initiate the data but processed and made sense of it.

*The machines existed because the information was already there.*

By the same principle, artificial intelligence can accumulate and process data and even enlarge its knowledge base. But it does so not by creating new bits of information but by gathering the information already present. The AI machines exist because the information exists.

Henley further writes:

Again, all of this makes even more stunning the Bible’s revelation that “in the beginning was the Logos,” and that “nothing was made that was made without the Logos” (John 1:1). Usually, the Greek *logos* is translated in the English Bible as “word,” but that term does not capture the scope of the meaning of the Greek *logos*. Heraclitus, for example, used *logos* to refer to “the ordering principles of the universe.” [Gra]

But what was it that had to be ordered, if not the innumerable bits of quanta or information? And what was the origin of the information?

Imagine the engineers building artificial intelligence machines. The technicians program the information into the box based on what they and others know in their own brains. Information is ultimately intelligence, and intelligence arises from minds with the capacities of reflection and reason, which means “person.”

Therefore, if information exists prior to creation, if information is intellectual data known in a mind, and if a reasoning mind is an attribute of person, then who is the Person who exists prior to the universe who possesses *logos* and is *logos* and infuses into the void the information that will bring the cosmos into existence?

The answer is God. Henley then goes on to say, “Remember that in the absence of a Creator, the skeptic would have to believe that the universe created itself, having first created the information by which it created itself.” And one final paragraph states:

Therefore, the AI phenomenon gives us a “backdoor” proof, or apologetic, of God’s existence. As the first AI machine could not have brought itself into existence but required a creator acting on the information in his or her mind, so the “vast quantum computer” that is the universe could not have brought itself into existence without something “transcendent” (other) to itself who could infuse it with information.

When the apostle John said what he said about the Word, he was saying something very profound: “the Word,” the information, had to come first. “In the beginning was the Word” means He existed in the beginning, and moreover “the Word was with God, and the Word was God.” He was in the beginning with God. God is immutable, pure, eternal being. He is not
becoming because He never changes, while all His creatures constantly change. Eternal God is just that: eternal. And He has a mind that is a reasonable mind and a revealing mind. He speaks. We see then in John 1:1 that everything that was made was made by the Word—the incarnate Word is the God who spoke everything into existence, as in Genesis 1.

Christ in Creation

So, looking at John 1:1–5 a little more closely, there are three attributes that God declares about the Word. Three basic realities about God are revealed in this section [Mac06]. First is His preexistence in verse 1: “In the beginning was the Word.” The phrase “in the beginning” refers to Genesis 1:1 where at creation the Word was already in existence. The eternal Son of God is not a created being. He was there in the beginning when everything was created (John 1:2). This describes eternal existence before creation—the eternal preexistence of the Son who is called the Word. Said differently, the Word didn’t become anything because He always was. The One who always was took on true humanity in His incarnation. But in His nature, He is pure, everlasting, and eternal being. That is why His name is I AM. In John 8:58 Jesus said, “Before Abraham was, I am,” and Colossians 1:17 says of Christ, “He is before all things, and in Him all things consist.” He was there with God in the beginning, and this declares His preexistence.

Second, John also points to His coexistence in verse 1: “He was God.” In His preexistence the Word was God. This is a very powerful Greek expression meaning “face to face with God,” which is to say He has intelligent personal fellowship on an equal level. Scripture affirms the deity of the Lord Jesus Christ. He is truly God, not an attribute of God or an emanation from God. He is not a creation of God, but He is God and at the same time a unique person who was with God. As Paul said to the Colossians, “In Him all the fullness of Deity dwells” (Colossians 2:9).

So the Word—the intelligent, divine, triune Being who created everything—necessarily preexisted His creation, because intelligence has to preexist the information required for the creation of anything. His preexistence is His eternality. Secondly, He is God, fully God, truly God, and with God. His coexistence is His equality. And thirdly in John 1:3–5 there is the declaration of his self-existence: “All things came into being through Him, and apart from Him nothing came into being that has come into being.” That’s true because in Him was life. He is self-existing—the uncreated One. “Jesus Christ is the same yesterday, today, and forever” (Hebrews 13:8). All things came into being because of Him. Notice that is a positive declaration. It is a simple, clear, and flawless proof of the eternal nature of the Lord Jesus as God. This doesn’t deny God as Creator, and it doesn’t deny the Spirit as Creator. The whole Trinity is involved in the creation. For further affirmation, a negative declaration follows the positive one: “Apart from Him nothing came into being that has come into being.” Not even one thing
exists that He didn’t create, and such is the only explanation that makes any sense because information must proceed creation. If in the beginning there is nothing, then “nothing” cannot generate the information needed to create. It is insane to deny that reality.

In verse 4 John also says of the Word: “In Him was life.” Life proceeded from Him. Everything that lives receives life from Him. This is a massive statement. “Life” here is the Greek word ζωή not bios. It refers to all kinds of life—not just biological life but life in every sense, both physical and spiritual. He has life in Himself. He didn’t receive life. He wasn’t given life. All life comes from Him. This self-existence is referred to by theologians as the aseity of God and of Christ. The creation of anything requires either a self-existent God, or a self-existent source of information without a mind and without a person. The latter is impossible. Out of nothing cannot come the information that creates everything because nothing is nothing. The truth, the reality, the information has to preexist the creation. Self-existent, eternal Deity spoke creation into existence. Again, notice that everything in the created world is becoming. But God is not becoming; God is being. He is pure, eternal being. Everything else is changing, and this distinguishes the creation from the Creator. Such is the fundamental reality of all reality: “In the beginning, God created the heavens and the earth.”

John further writes in verse 4 of chapter 1, “the life was the Light of men.” While some distinction obviously exists between life and light, here they are inseparable: life becomes the Light. The life is the manifestation of divine truth. The Light displaced the dark void with truth. God’s life shines like holy light into the darkness of sin, and “the darkness did not comprehend it” (verse 5). Literally, it did not overpower it. And we know the darkness did not overpower it because we have received the Scripture—the revelation of God as to the truth of the life and the Light.

Summary

So the role that Christ played in creation is stated directly in John 1:1–5. By just touching lightly on this text you recognize that it affirms that there must be information before creation. Information means intelligence, intelligence means a mind, a mind means a person, and a person has to be someone who is preexisting—someone who is transcendent. That points to God and God alone.

There is a reason why Genesis 1 is under attack: if God can be eliminated at the very opening of Scripture, then rejectors have reason to discredit all the rest. So, John gives us an unparalleled look in just those few verses at not only the deity of Christ but also the fact that He is the divine revelation that brought everything into existence. And in the revelation of God in creation and in the spiritual realm, He has provided the Light of salvation that extinguishes the darkness. So the creation account cannot be separated from the gospel. The Light of men is the gospel. The Light shining in the darkness is the light shining against the
blackness of sin. The same One who created everything is the One who shines gospel light into the darkness.

You wouldn’t want to tamper with Genesis 1, nor would you want to tamper with John 1. These two accounts give us all we need to know. God created by speaking. Christ, the Word, created everything that existed by speaking. He spoke, and life came into existence—including spiritual life, which comes like light in the darkness to deliver men from sin and death and judgment.

Bibliography


