ASTERIA

ISSUE 1 • --/--/2022

Contents

Foreword

Welcome to Asteria!

From a column on the inner workings of human physiology to critiques of renditions of Chopin, this is a safe platform for students to come forward and share their passions in academia.

Throughout the production of this initiative, the opportunity to work with so many amazing personalities is a truly rewarding experience. The best part? We're just getting started.

Our mission is to build collegial partnerships with the secondary schools we work with. In time, we hope to grow a community around this initiative.

This project would be impossible without the contributions of Arya, Mudit, Harry, and our editors, who have dedicated their time and energy to present you with the breadth of knowledge on offer from our excellent writers.

A big thank you to Reading and Kendrick School for their ongoing cooperation in this student-led initiative.

If you would like to contribute or join us, feel free to contact myself or my peers.

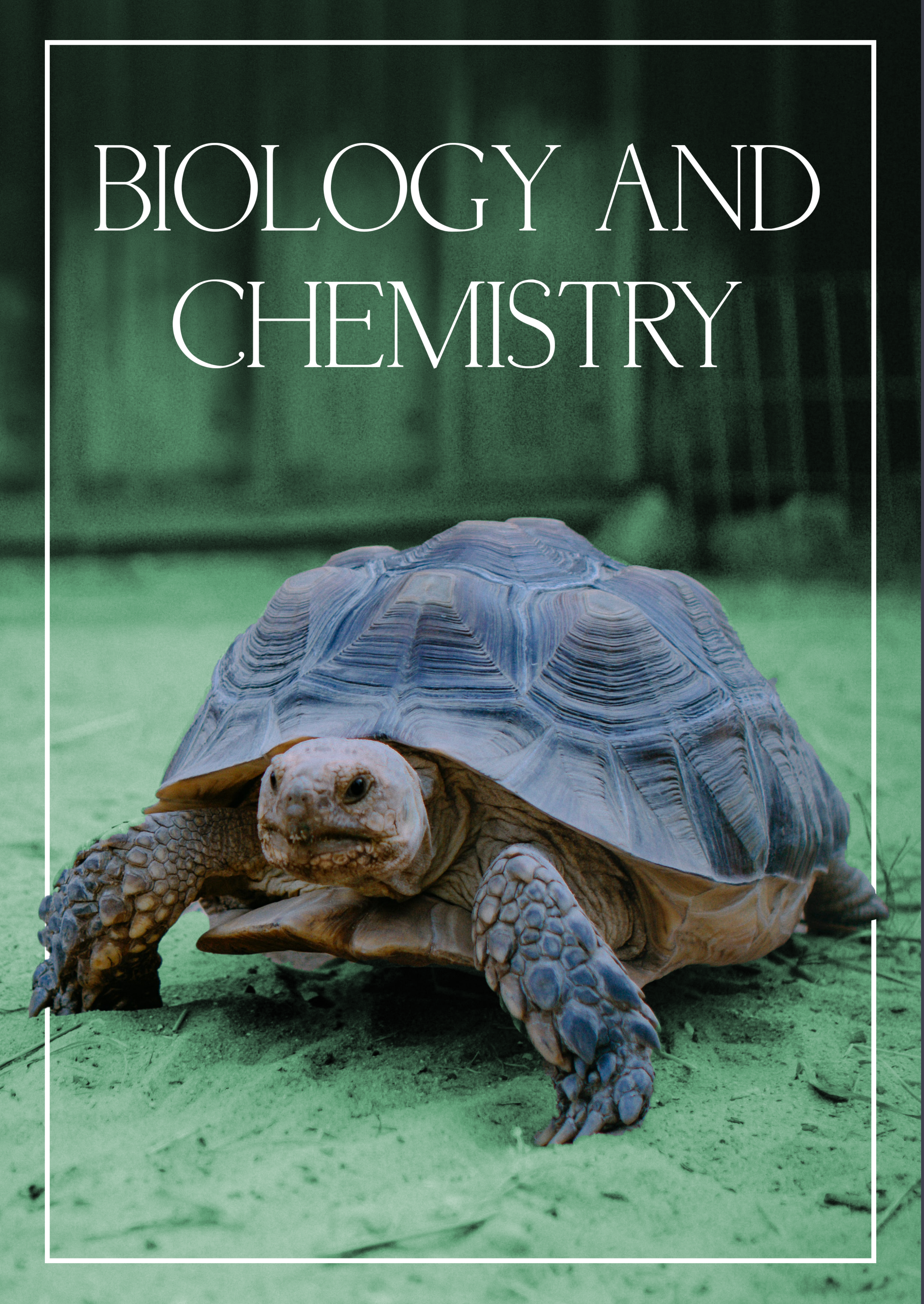
Ram Gautham Kurakula

Ram Gautham Kurakula
Chief editor, Co-ordinator

Arya Kalavath
Editor, Co-ordinator

Mudit Pahwa
Designer, Co-ordinator

BIOLOGY AND CHEMISTRY



Electrical signalling in the body

Anoushka Barretto

Harvey Zhang

Whether neurons are sensory, motor, big or small, one thing they all have in common is that their activity is both electrical and chemical. The cooperation between neurons forms the basis of our central nervous system and understanding how cells conduct electricity is vital to grasping its complexity, as well as aiding us when problems arise.

In this article, we will be discussing the fundamentals of electrical signalling in the body. Electricity is required for the nervous system to send signals throughout the body and to the brain, making it possible for us to move, think and feel. This is only possible through our nervous systems and the use of neurons and action potentials.

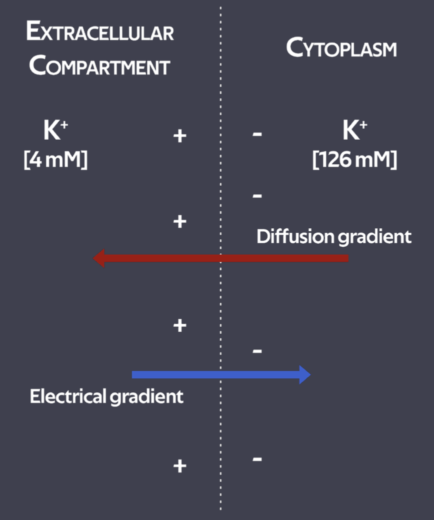
In the early nineteenth century, the nature of electricity remained vague to scientists. The history of bioelectricity began in 18th century Bologna, Italy when scientist Luigi Galvani was dissecting a frog[1]. As his assistant touched a nerve in the frog's leg with a metal scalpel, he noticed that it picked up static charge causing the frog's leg to move. Later another scientist called Alessandro Volta repeated Galvani's experiments[2], but doubted the existence of an electric fluid intrinsic to animals. Further research led to him developing the first battery, proving that electricity could be generated outside of a living creature. Today, galvanism is known as the electrical phenomena in living beings.

Firstly, in order to understand the basics we have to be aware of what a neuron is. A typical neuron contains a cell body and some which houses the cell's nucleus and a variety of cellular machinery. It also consists of axons and dendrites extending out as well as an outer layer which is composed of a bilayer of lipid molecules that partition the inside of a cell from the outside. This is known as the cell membrane.

Let's start with the resting potential. When a cell is at rest, it is not electrically neutral. A small voltage is measured known as the membrane potential, which calculates to around 70 millivolts. The membrane potential of a neuron that is specifically "at rest" means that it is not sending or receiving signals. Another term we should be familiar with is the membrane potential. This is a general term that describes the voltage across the membrane. At any point in time, the membrane potential of a neuron can vary widely for example from -90mv to +60mv.

Equilibrium potentials

Now that we have visited resting and membrane potentials we shall introduce equilibrium potentials. The equilibrium potential is the electrical potential difference across the cell membrane that exactly balances the concentration gradient for an ion. Consider the case for a single ion. Movement of charged ions through a semipermeable membrane is governed by two forces: the concentration gradient (a difference in concentration) and the electrostatic gradient (an electrical potential difference). As ions move, the concentration gradient that drives movement decreases while the electrostatic gradient that opposes movement increases. This means that an equilibrium will always be reached where the concentration gradient is exactly equal and opposite to the electrostatic gradient. At this point, the electrostatic gradient is known as the equilibrium potential.



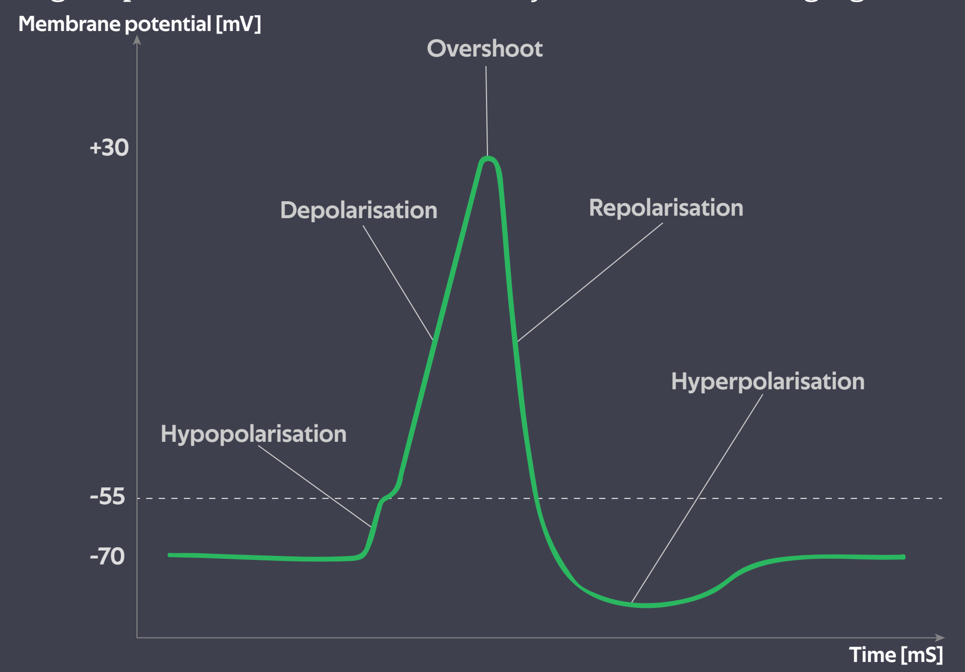
This figure shows the movements of potassium ions across a cellular membrane- potassium ions leave the cell according to its concentration gradient, however this is accompanied by the movement of potassium ions into the cell as they are attracted towards the increasingly negative intracellular environment[3].

Several factors influence this equilibrium potential: the concentration of ions on both sides of the membrane (concentration gradient), temperature (increases the effect of the concentration gradient), and the charge of ions (increases the effect of voltage). To maintain stable concentrations of both sodium and potassium, a sodium potassium pump is essential to maintain high intracellular potassium and high extracellular sodium. This pump actively transports potassium and sodium against their concentration gradients; this allows the concentration gradient that these ions travel down to be maintained and therefore, for the resting membrane potential to be maintained. You may have heard of the sodium potassium pump from your studies in Biology at GCSE and A Level, an example of active transport.

The action potential

Now that we have visited equilibrium potentials, it is time to consider action potentials. Action potentials occur when a neuron sends information down an axon away from the cell body. Neuroscientists use words such as spikes or impulses for the action potential as it is known as an explosion of electrical activity that is created by a depolarising current.

Embedded in the membranes, channels change in response to their local environment, opening and closing at any given moment depending on their local membrane potential. Complex molecular machines are made from proteins. As voltage changes, the conformation of the protein shifts either opening and closing the pore of the channel, hence they are known as voltage-gated.



This figure[4] shows the stages of the action potential, explained below

The action potential can be broken down into a few stages. The first stage begins with the activation of the action potential, during which a few Na+ channels open causing depolarisation to reach threshold potential (the potential when the inward Na+ current balances the outward K+ current). When this threshold potential is exceeded, an action potential can be generated. The second stage is when depolarisation is sensed by voltage gated Na+ channels which activate and allow Na+ in, causing the cell to depolarise further. Therefore, a positive feedback loop is set up as depolarisation leads to more voltage gated Na+ channels opening, leading to more depolarisation.

Following this, all Na+ channels open and the membrane potential approaches ENa; the membrane becomes more permeable to Na than K. At 40mv, Na+ channels shut automatically. It is important to realise that during this period, K+ channels have been slowly opening, allowing the positive charge out. K+ channels are voltage gated too but open slowly in response to depolarisation. After this, it follows that repolarisation occurs as K+ channels allow K+ efflux and Na+ channels are shut, bringing the membrane potential back towards Ek. Finally, some K+ channels close and the membrane stabilises at the resting membrane potential.

However, the potential overshoots the resting level because K+ channels are also slow to close again on repolarisation. This is the hyperpolarisation period as the membrane potential is more negative than the resting potential. So, what's happening is that a negative feedback loop is taking place: depolarisation leads to an increase in permeability to K+ leading to an increase in K+ out, repolarising to counter depolarisation.

Propagation of the action potential

How are action potentials propagated? Action potentials are propagated by local currents along the membrane that change the voltage of neighbouring regions. When depolarisation occurs at one region of the axon, the current generated travels passively along the axon. Using our knowledge of the action potential, this spread of current depolarises adjacent regions of the membrane, opening some Na+ channels in the adjacent membrane. If the current is large enough, the depolarisation of the adjacent region will be large enough to exceed threshold potential. A positive feedback loop is set up leading to full depolarisation and initiating an action potential in the adjacent membrane regions. The depolarisation in turn generates currents which spread and depolarise the neighbouring region of membrane, and the cycle repeats. Eventually the signal travels down the axon and generates an action potential at the opposite end of the axon.

Another important question to ask is how are action potentials propagated in one direction? The absolute refractory period occurs when Na+ channels are inactivated after they have just closed and will not open regardless of membrane potential and depolarisation, the Na+ channel will return to its active state after a period of time, ready to be activated when the membrane is depolarised. The relative refractory period occurs when the threshold potential is raised due to the reduced availability of active Na+ channels. Only some Na+ channels can be activated, thus a greater depolarisation is needed to generate a large enough inward Na+ current to balance the outward K+ current for threshold potential.

These refractory periods allow the conduction of action potential in a singular direction. An action potential cannot be propagated backwards, only forwards. When an action potential is generated, the current generated can cause depolarisation of both the upstream and downstream adjacent regions. However as the upstream region will still be in a refractory period it is impossible or more difficult for an action potential to be generated due to the lack of Na+ channels available for activation.

Similar in effect to the relative refractory period, K+ channels may still be open from the previous action potential in the upstream region. Thus the outward K+ current opposes the depolarisation experienced, reducing the effect of the Na+ current on membrane potential. The membrane upstream may still be hyperpolarised, so a much larger current is needed to bring the membrane potential up to the threshold potential.

Overall the action potential generated is much more likely to only initiate another one in the downstream region where it is not in refractory period and the membrane is at resting potential. Na+ channels are active and ready to be activated. In this way, there is a one way propagation of the action potential down the axon!

The length constant

An important consideration when it comes to action potentials is the length constant. The axoplasm has a high resistance, longitudinal resistance along the axon is too high. In contrast, cell membranes are poor resistors, and as a result transmembrane resistance across the membrane is too low. The result is that electrical current generated during an action potential leaks out of the axon as it travels through the axoplasm.This means voltage decays with distance from the action potential where initial depolarisation occurred. We define the length constant as the distance needed for the voltage to decrease to 37% of the initial voltage from the action potential. This describes the extent of leakage. In human axons the length constants are very small. Therefore the passive spread of current is unable to carry information over large distances.

This is the very reason why action potentials exist and need to be constantly generated along the axon to allow for constant injection of currents so the signal (an action potential) can travel from one end of the axon to the other. An action potential also has to be propagated by a series of currents!

How can we speed up propagation of an action potential?

What affects the speed of propagation of an action potential? One logical way of speeding up propagation and making action potentials faster would be increasing the density of sodium channels. More channels represents a lower resistance and more current. But this is metabolically expensive as more sodium will need to be pumped out of the cell to active the action potential. What about increasing the length constant? This would allow currents to spread further along the axon; regions further from the active regions can be brought to threshold by passive current flow, speeding up the action potential. In theory, this could be done by decreasing longitudinal resistance or increasing transmembrane resistance. However, the things that determine longitudinal resistance (width and length of nerve and the intrinsic ionic composition of the nerve fibre itself) are not easy to change and are already optimised.

A particularly well known adaptation of neurons is myelination- axons are wrapped repeatedly around by schwann cells (in PNS) or by oligodendrocytes (in CNS), covering the axons with concentric layers of myelin. The myelin sheath increases transmembrane resistance, increasing the length constant. Thus less current can leak through the membrane and voltage decays less rapidly. It is important to note that myelination increases length constant but does not change the intrinsic longitudinal resistance of the nerve itself.

An additional note is that small-diameter myelinated nerve fibres conduct action potentials slower than unmyelinated fibres. This is because the axon of the myelinated fibre is too narrow causing the longitudinal resistance to be too great to conduct action potentials effectively. Hence small nerve fibres e.g. pain fibres are unmyelinated.

Naziyah Hasan

Glaucoma: the leading cause of blindness

I – Introduction

Recent advances in medicine have led to the development of numerous devices that aid the detection of eye conditions such as glaucoma, macular degeneration and cataracts which are amongst the causes of reduced vision or even irreversible blindness. With age as the largest risk factor for most of the aforementioned conditions, screening the elderly at higher risk has become a common practice in the UK, allowing early detection and thus reducing complications or blindness.

As mentioned by Glaucoma UK[1], glaucoma is the leading cause of blindness affecting over 700, 000 people in the UK of all ages. It is a chronic, neurodegenerative disease and is defined as the loss of retinal ganglion cells in the optic nerve of the eye, resulting in progressive loss of peripheral vision. The damage to the optic nerve head is due to elevated intraocular pressure (IOP) as a result of unevenness between the production of aqueous humour (transparent fluid located in the eye) and the drainage system of the eye.

II – Development of glaucoma

The aqueous humour, as shown in Figure 1, provides nutrients to the eye and maintains the eye pressure. It is produced by the ciliary body and flows to the anterior chamber where it is drained at the drainage angle. The reference of 'angle' is the angle between the cornea and iris. Amongst the various structures of the drainage angle, the most essential structure in the drainage of aqueous humour is the trabecular meshwork which is a spongy tissue with a sieve-like structure.

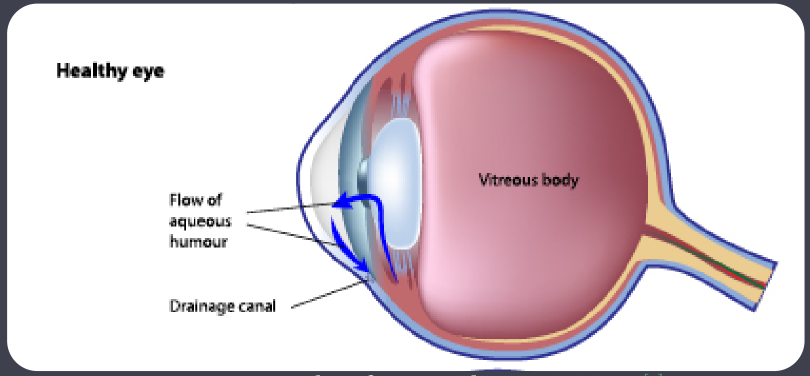


Figure 1 – taken from Medirex Opticians[8]

Following fluid drainage from the eye via the trabecular meshwork, the fluid flows through a structure called the Schlemm's canal, into collector channels, veins and then eventually, back into the body's circulatory system[3].

A blockage in the drainage system restricts the flow of aqueous humour into the anterior chamber, resulting in increased pressure in the eye and eventually damaging the optic nerve. The optic nerve (also known as second cranial nerve) found at the back of the eye transfers visual information from the retina to the brain via electrical impulses[2]. The high pressure caused by inefficient drainage compresses the second cranial nerve resulting in cell death – this is called Glaucomatous optic atrophy. If left untreated or undetected, this can lead to irreversible blindness.

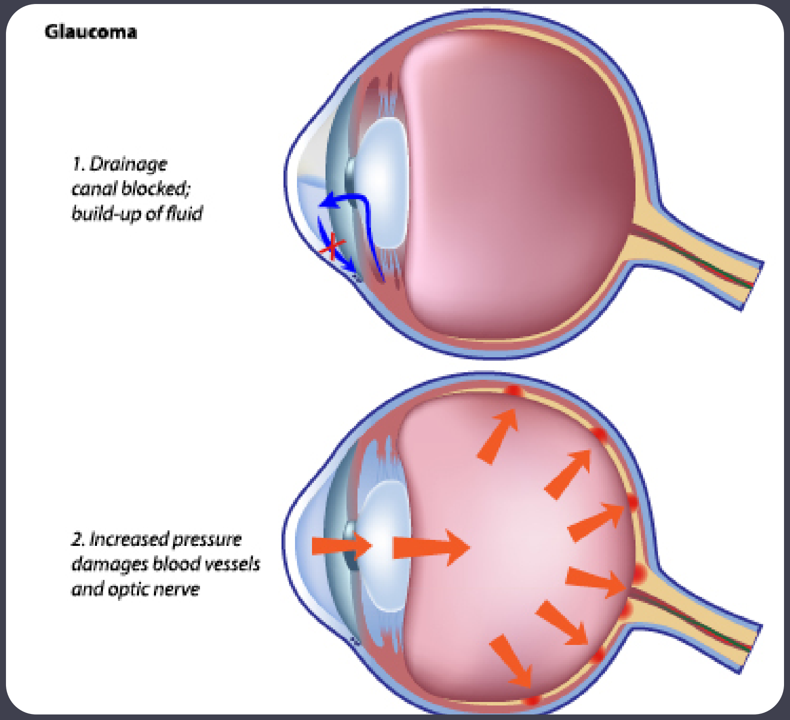


Figure 2 – taken from Medirex Opticians[8]

Recent research in Singapore has led to the discovery of a gene mutation linked to exfoliation syndrome (XFS), which is known to be the basis of secondary glaucoma[7]. If an individual has XFS, fibre-like substances existing outside of cells are deposited on ocular structures and throughout the body. The exact reason for XFS is unknown and hence, the mystery surrounding it means that the ultimate cause of glaucoma is still to be concluded.

III – Classifications of Glaucoma

There are several types of glaucoma, some include:

- Juvenile/adult Primary open-angle glaucoma
- Acute-angle closure glaucoma
- Secondary glaucoma
- Congenital glaucoma

The most common type of glaucoma is primary open angle glaucoma (POAG). Anyone can develop POAG but there are some risk factors that may make individuals more susceptible to it. These include:

- Age
  - According to Glaucoma UK[6], it is estimated that about 2% of people aged 40+ have POAG, and this rises to almost 10% in people older than 75 (Glaucoma UK, n.d.)
- Ethnicity
  - Individuals with an African-Caribbean origin are at a higher risk of developing POAG compared to an individual with a European origin
- Family History
  - If your parents or sibling suffer from glaucoma, you are at a higher risk of developing POAG
- Myopia
- Corneal thickness

Like most types of glaucoma, fluid is unable to be drained which leads to an increased eye pressure, damaging the second cranial nerve. Nevertheless, in POAG, a subgroup known as normal tension glaucoma (NTG), where the eye pressure is not very high, still damages the second cranial nerve. This may also be referred to as low tension glaucoma. Unlike most eye diseases, glaucoma does not show symptoms in early stages as there is no pain or drastic change in vision since the damage occurs in the peripheral parts of the vision.

As mentioned, glaucoma left undetected can lead to devastating consequences such as irreversible blindness. So how can glaucoma be detected?

IV – Detection of Glaucoma

After years of research and developments, there are now several modern computer-based machines that can be used to detect early signs of glaucoma. Some include the optical coherence tomography (OCT) machine and the fundoscopy machine.

The OCT machine is a high-resolution imaging modality producing a cross-sectional tomographic (3-dimensional) image of the retina and evaluates the optic nerve. The use of long-wavelength and broad-bandwidth light sources to illuminate the retina enables it to assess the light reflected from the retinal tissues through the use of a spectrometer[4]. Additionally, it can indicate small changes in the thickness of the nerve fibre layers which gives early signs of glaucoma. Other than glaucoma, the OCT machine can also detect age-related macular degeneration (AMD), diabetic macular oedema (DME), central serous chorioretinopathy (CSR) and many more diseases. A low image quality can often be caused by cataracts due to the blockage of light.

In addition to the OCT machine, a fundoscopy machine can also be used to detect glaucoma. Like the OCT machine, it is a modern medical imaging technique that enables optometrists or ophthalmologists to detect changes in the optic disc and hence, glaucoma.

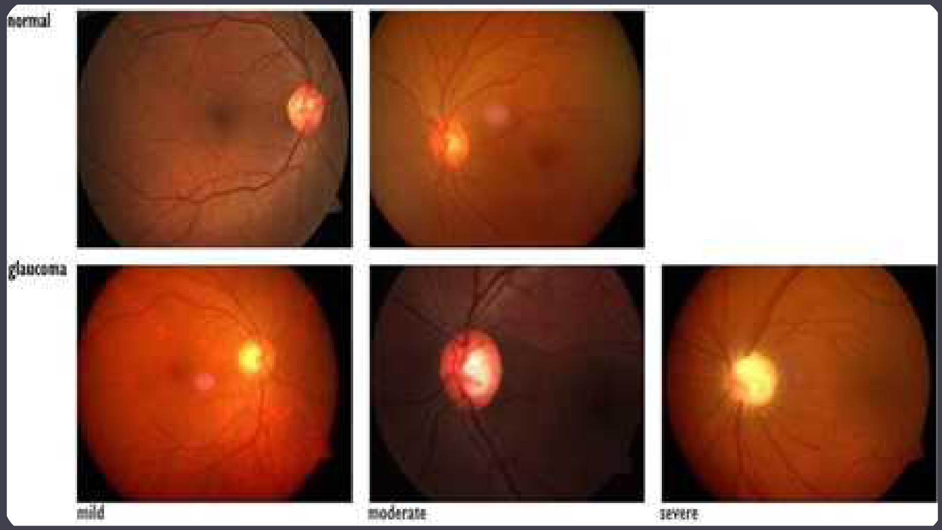


Figure 3 – fundoscopy pictures from Hindawi Journal comparing optic disc without and with glaucoma[5]

V – Treatment of Glaucoma

Although there are no available treatments to regain vision that has been lost, eye drops such as betaxolol hydrochloride (Betoptic S) and metipranolol (Optipranolol) have been made available to those suffering from glaucoma. These are beta blockers which reduces the production of fluid released from the eye. Prostaglandins such as latanoprost (Xalatan) are other forms of eye drops that improve the drainage of the fluid in the eye to reduce ocular hypertension. Other forms of treatments include laser procedures and surgery. Nevertheless, there is still room for further developments of treatments such as gene therapies to prevent more prevalent ocular diseases such as POAG and its eventual testing in humans. Despite ongoing research on how gene therapies may prevent POAG, Luxturna, a form of gene therapy, has been known to treat congenital blindness (another classification of glaucoma) which is a seminal accomplishment.

VI – Conclusion

Having said that, it is important to remember that these non-invasive treatments do not cure glaucoma, but instead it controls the intraocular pressure in the eye and prevents damage to the optic nerve. Hence it is vital to regularly attend eye check-ups, especially with a family history of glaucoma or other eye conditions.

Don't forget: Prevention is better than cure.

Water intoxication: an unlikely poision

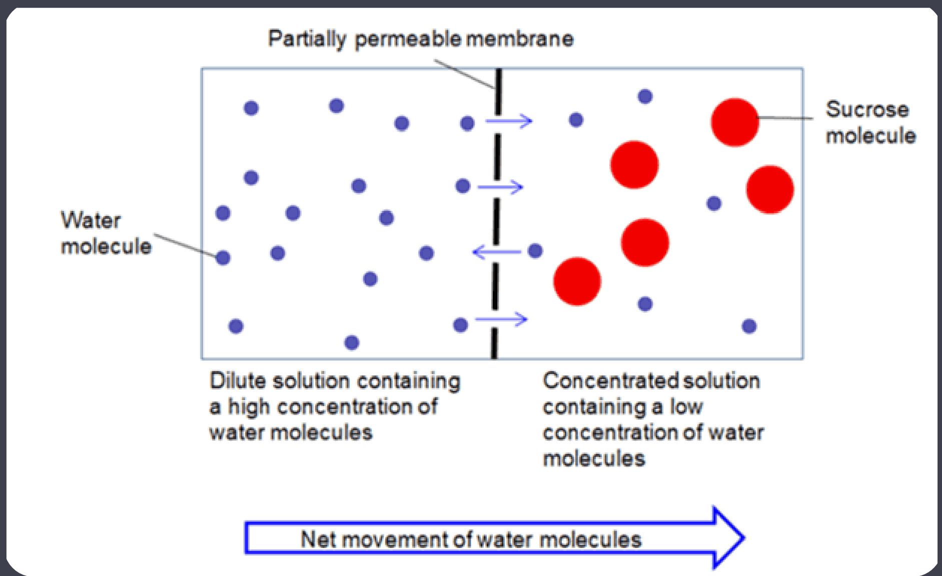
Mathew Wong

12 January 2007. California. On a morning radio show, an excruciating contest was going on. Named 'Hold your wee for a Wii'[1][3] by the now-defunct KDND radio station, 18 contestants were downing bottles of water one after the other, on the condition that whoever could drink the most while not urinating would win a Nintendo Wii gaming console as a prize. Contestants agonizingly crammed down mouthfuls of liquid down their throats, while the hosts egged them on with gleeful laughter. As time ticked by, some could no longer endure such an ordeal, dropping out in the middle and probably wondering why they had signed up to such a torturous session. "She looks like she's pregnant," sniggered one of the DJs, pointing to a female participant with an overwhelmingly bloated stomach. Little did they know, she would be dead in a few hours.

Jennifer Strange, 28, had passed away from overhydration, by literally drinking much more water than her body could handle.

While it's "common sense" that the overconsumption of pretty much any substance can lead to some unpleasant repercussions, the exact science behind it isn't so straightforward. In this article, I will attempt to break down the mechanism of water poisoning in simple layman's terms.

When excessive amounts of water are consumed in a short period of time, the kidneys, the organs responsible for filtering waste out of your body fluids, simply cannot work fast enough. Liters of water are now circulating throughout the bloodstream, diluting blood to a large extent. This means that there is a very high concentration of water molecules outside of the cells. In contrast, the fluid inside human cells consists of a wide array of substances, with potassium, magnesium and sulfate ions just to name a few. Hence, at this point, the concentration of water molecules is lower inside the cell than outside (it isn't as diluted)

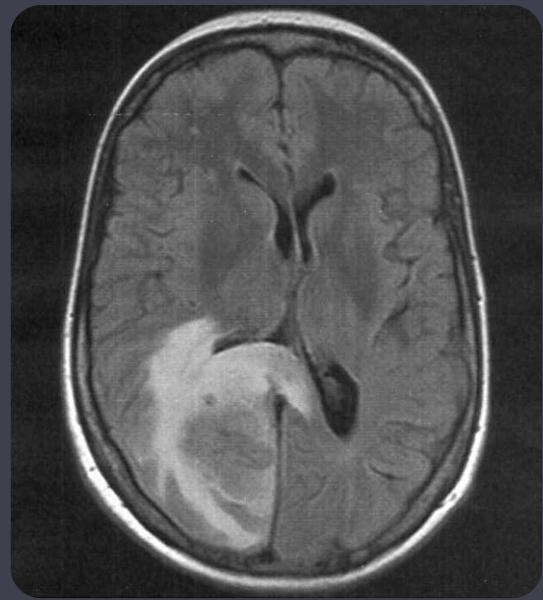


A diagram that demonstrates the process of osmosis, using sucrose and water as examples.

And this is when the principle of osmosis comes into play. For those who may not be aware, osmosis is when solvent molecules naturally diffuse from an area of high to low concentration across a semi-permeable membrane. For the sake of argument, let's say that across such a membrane there are two sides, one with 100% water and the other side with 50% water. In this case, we would see water moving into the side with less water molecules (i.e., the 50% side). To visualize this phenomenon easily, try placing a dried raisin into water. Overtime, it will slowly expand in size as it absorbs water from its surroundings. Why? Because the inside of a raisin has less water than the outside, and by osmosis

water will go in. The same applies to human cells. With much more water present in blood, they are taken up by cells which swell up massively to accommodate more liquid.

Swelling cells isn't that big of an issue for many parts of your body, but that doesn't apply to the brain. Enclosed tightly inside the skull cavity, brain cells simply have no room to expand. What then occurs is the swollen brain tightly pressing against the sides of the skull, leading to headaches, one of the first symptoms to be observed, accompanied with nausea and mental disorientation. The heavy pressure also obstructs blood flow, leading to chances of seizures or the victim simply falling unconscious. This phenomenon, cerebral edema, is known to be a major cause of strokes and traumatic brain injuries.

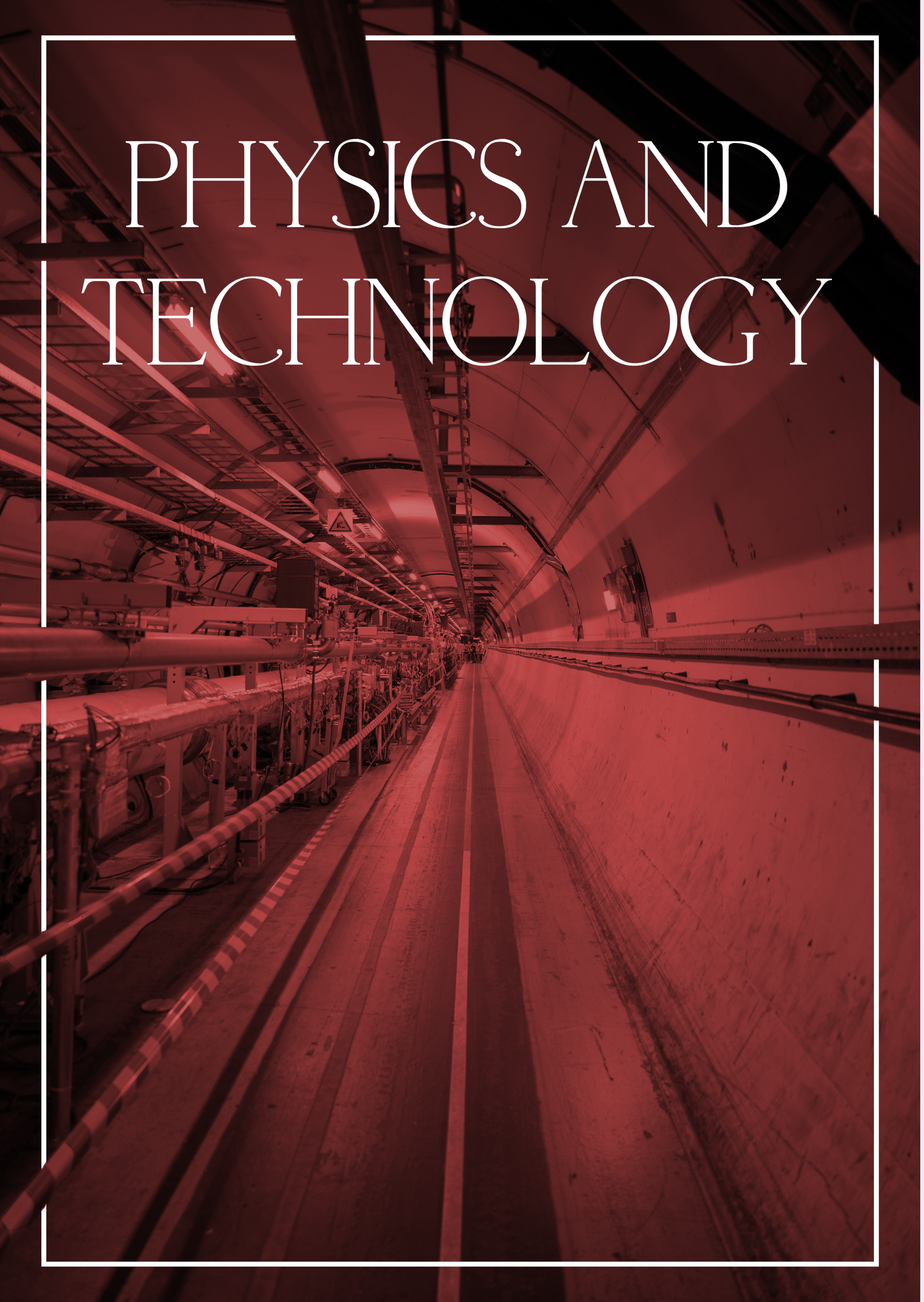


MRI Scan of a brain affected by cerebral edema[2]

One prominent risk factor that increases the risk of overhydration is heavy sweating and physical activity, when one is more prone to drinking an excessive amount of water unknowingly. Marathon runners are particularly susceptible, with a 2005 study[4] showing that 13% of Boston Marathon participants ended up with hyponatremia (low sodium concentration in blood), a symptom of overhydration. The researchers concluded that this was linked to "consumption of more than 3 liters of fluid during the race". In a separate investigation by the American armed forces[5], three soldier deaths attributed to water intoxication were uncovered, where the water intakes of the victims "reached 5 liters, usually 10-20 liters in just a few hours".

And what can be done about this? In serious cases, diuretics, a class of drugs that increases urine production, can be administered to get rid of excess water quicker. A second treatment involves the use of VRAs, substances that interfere with the hormone vasopressin at its receptors. This hormone instructs the kidney to conserve more water, and by inhibiting its action, urine volume is increased.

If you're feeling ever so slightly uneasy having read to this point, please do not be concerned. Healthy human kidneys can process around 1L of water (Approx. 4 average-sized glasses) per hour, and in many cases the condition is relatively mild where no significant symptoms are noticed. Considering that the commonly said "8 glasses of water per day" only amounts to around 2-3L, there should be zero concern for overhydration as long as you're not guzzling more than half of your daily water intake in a single sitting.



The enigma of Hoag's Object

Harsha Kuruganti

Introduction

Almost every galaxy in our universe can be categorised into three types: Spiral, elliptical, or irregular (see Fig.1). This system of classifying galaxies via their shapes was first established by Edwin Hubble[1].

Spiral galaxies are further divided into three classes based on their shapes: Ordinary spiral galaxies (consisting of a central bulge and spiral arms), barred spiral galaxies (featuring an elongated bar of stars across the central bulge), and lenticular spiral galaxies (having disc-shaped features as opposed to spiral arms)[1]. These galaxies appear to be very "dusty" and make up approximately 60% of all galaxies.

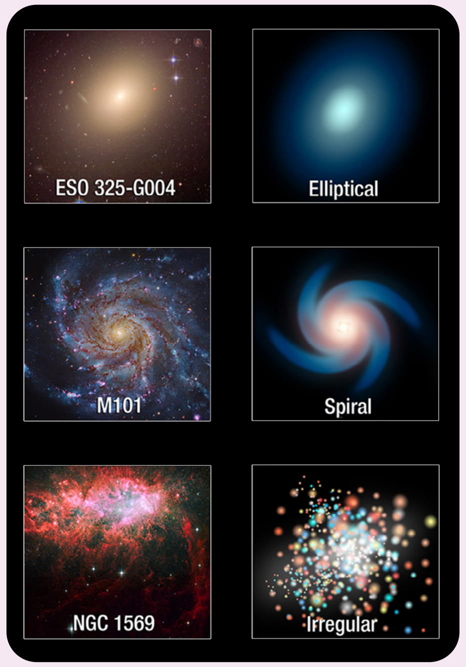


Fig. 1 – Illustrations of the three main types of galaxies (right) with actual photos of galaxies (left) that fit the categories[2]

Elliptical galaxies, on the other hand, are differentiated based on a scale ranging from E0 to E5, where E0 is spherical and E5 is elongated. They are observed to have very little dust and gas compared to spiral galaxies, but appear to be older[1], making up 10-15% of all galaxies.

Lastly, irregular galaxies consist of galaxies that have no regular shape or form. They are estimated to make up 25% of all galaxies, and are believed to have been more prevalent in the early universe. Additionally, these galaxies are generally smaller, and are often satellite galaxies of larger spiral or elliptical galaxies[1].

However, some galaxies fall outside this classification system; approximately one in ten thousand galaxies fall into one of the rarest categories – ring galaxies.

Fig. 2 – the galaxy NGC 6028 representing an ideal ring galaxy[3]

Properties of ring galaxies

So, what are some key features of these phenomenal galaxies? Starting with the most obvious property, a ring galaxy consists of a central core surrounded by a ring of luminous stars. However, between these layers, there exists a gap (shown in Fig. 2); a region of low density. Almost no stars and very little gas separates these two layers. The relatively compact core has a low gas composition and consists primarily of older stars[3]. Comparatively, the ring appears to be "bluer" than the core (due to the stronger concentration of short-lived blue stars), indicating that the stars within the ring were formed more recently. Additionally, when astronomers observed where these galaxies were generally located, they found that the galaxies were spread out in a region commonly referred to as "the field"[3] as opposed to the more central location of galaxy clusters. Recently, in 2018, Elena Bannikova, a professor of astronomy at the University of Kharkiv, analysed the motion of particles within a specific ring galaxy – the Hoag's Object – using a theoretical and mathematical approach. From her calculations, she found that the forces between the ring and the central core balance on

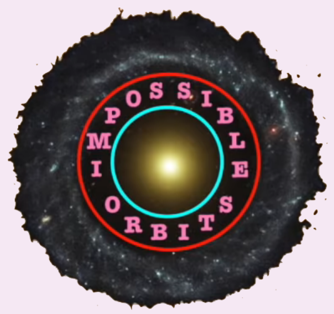


Fig. 3 A graphical representation of the point at which the forces balance (red) and the last circular orbit (blue)[4]

the inner edge of the ring (shown in red in Fig. 3). She also calculated where the last stable circular orbit could occur (since the circulating object may be disturbed by the ring beyond a certain limit), finding a circular locus outside the core (see Fig.3). Between these two circles, there exists a space where circular orbits cannot exist. From these observations, scientists have been given an understanding towards why this gap between the ring and the core exists.

A brief history of ring galaxies

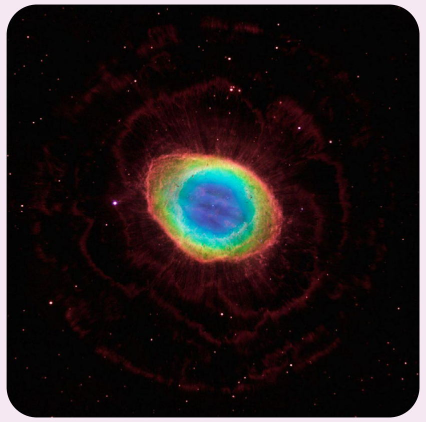


Fig. 4 – a composite image of the Ring Nebula[5]

Dating back to 1950, the first ever ring galaxy, Hoag's Object, was discovered by Art Hoag, baffling astronomers of its existence. At that time, the only method to making a ring-like object in space was where a central bright core was surrounded by a luminous halo i.e., a planetary nebula, such as the Ring Nebula in Fig. 4.

Over the years, the number of ring galaxies being discovered has multiplied drastically, especially recently, with the current number having increased to approximately 40,000. A project called "Galaxy Zoo"[6] recruited thousands of volunteers to identify the morphology of various galaxies[6], from which an AI algorithm, called "Zoobot" was produced. It utilised the inputs from these volunteers, and accurately identified galaxy shapes as well as predicting where it may be mistaken. This algorithm helped identify these ring galaxies, providing a dataset that can be used to reveal how galaxies evolve.

Hoag's Object, the first ring galaxy, discovered in 1950[3]

Formation of Hoag's object

There are several theories aiming to resolve the enigma of the formation of Hoag's Object, some more elaborate than others.

As mentioned earlier, Hoag's Object (Fig.5) was first discovered in 1950, named after astronomer Arthur Hoag. When he first discovered this phenomenon, he explained that the "size and nucleus" were not of the typical planetary nebula, subsequently, labelling the object as a potential "new species" of galaxy[7]. Although he proposed other possible theories, including a diffraction effect or a gravitational lens system (occurs when a massive celestial object "bends" the path of light around it), he referred to them as "less probable explanations".

Looking deeper into the structure of the object, the halo was observed to consist of young, blue stars, as opposed to the strong emission lines expected by a nebulae model, proving that Hoag's object was of the galactic nature[3]. Furthermore, the idea of a gravitational lens or a diffraction phenomenon were both invalidated since the centre of the galaxy and the halo displayed identical redshifts, showing that they are both part of the same system[5]. So how is this galaxy formed?

One theory includes the unstable bar model, proposed by Noah Brosch in 1985[8], where the rotational velocity of the bar in a barred spiral galaxy increases to a certain point, until the bar structure becomes unstable, leading to a "spiral spin out". This would result in a ring-like structure. The creation of the luminous halo of stars is due to the density waves from the unstable bar spreading out into a ring, also called a resonance ring[8]. As a result, the increase in pressure and gravitational force caused by the density waves promotes star formation, leaving behind a centre of older stars. However, in 1987, Schweizer et al. provided a strong counter argument against this theory; they detected a H I component (which informs us about the different processes that shape a galaxy) much larger than the upper limit within Brosch's model[8], determining that the galaxy was in fact spheroidal, not disc-like which would be required for a bar instability to occur, disapproving of the Brosch hypothesis. Although it would be theoretically possible for a core to evolve from a disc shape into a spherical, they concluded that this scenario was unlikely[9]. Additionally, Brosch failed to detect remnants of a bar-like structure, further providing evidence against this theory.

Another theory proposed was that of a collision between two galaxies. Schweizer also discussed this

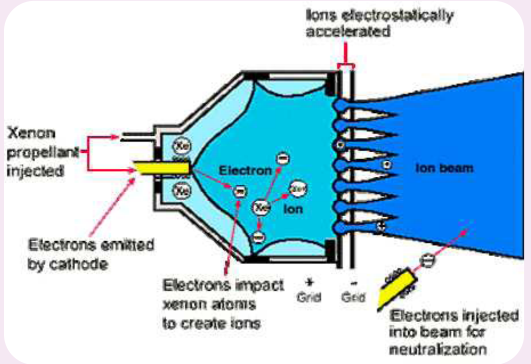
Space travel: the rise of a new age

Arnav Tyagi

July 16th, 1969. Apollo 11 had just launched from Earth, and was carrying on board the first people to ever land on the Moon. It was a pivotal point for mankind, and marked as one of the greatest achievements of space travel, paving the way for many more accomplishments. Over 50 years later, we stand once more, on a pivotal point like no other. We are on the cusp of being able to colonise Mars, discover the secrets of new galaxies, and commercialise space travel for all. Let's have a look at some of the ground-breaking ideas that will help us make these dreams reality.

Firstly, let's look at rockets. We currently use chemical thrusters to reach space, combusting liquid oxygen with liquid hydrogen. Although it gets the job done, it is extremely inefficient, as most of the weight of the rocket consists of fuel, and only 5%[1] of it is the payload- the object we want to actually send to space.

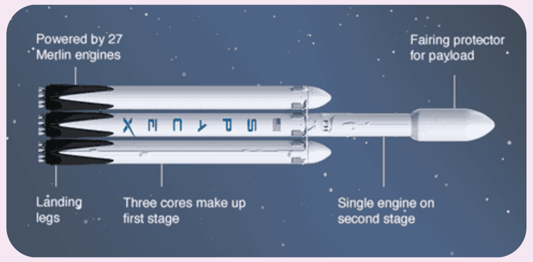
Fortunately, the future of rockets looks very bright, with lots of alternative fuel sources currently being proposed and researched into. As of now, the most successful projects are ones which use ion fuelled engines, better known as electric system engines, as they only use electricity to propel ions. Rather than lighting up liquid fuel to accelerate, ion fuelled engines spew ions at much faster velocities, and are significantly more efficient and longer lasting than chemical engines. The best chemical rockets have a fuel efficiency of around 35%, whereas ion rockets have an efficiency of 80-90%.[2]



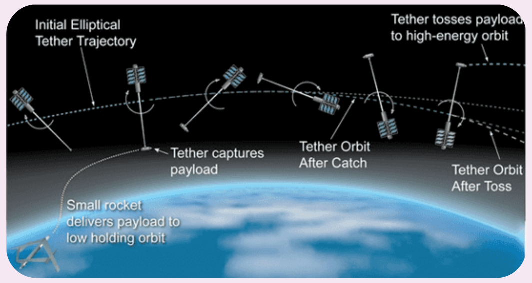
In an ion thruster, an inert gas (usually xenon) is bombarded by high-speed electrons inside a discharge chamber. These high-speed electrons collide with and strip away the electrons from the xenon atoms, creating more free electrons as well as ionising xenon atoms into positively charged xenon ions (this is called an ionising event). This soup of positively charged xenon ions and negatively charged electrons is now in a state called plasma. Although this plasma has both positive ions and negative electrons, overall, the charge is neutral. This is known as quasi-neutral plasma. Electromagnetic fields and a high voltage are used to speed up and push out the ions from the discharge chamber and out of the thrusters at very high velocities, accelerating the spacecraft. Once out of the thrusters, an electron beam shoots out electrons in order to neutralise the xenon ions, so they don't strip electrons from the spacecraft itself. The electromagnetic fields keep the extra electrons within the discharge chamber, increasing the ionising events, making it more efficient. They also minimise the electrons hitting the walls of the chamber in order to increase the engine's durability. This whole system is powered by solar panels, so no battery or power system is needed, minimising the total weight of the spacecraft.

Although these ion thrusters don't produce as much thrust as chemical rockets, and can't yet be used to escape Earth's gravity well, they can last for months at full throttle (until they run out of fuel). This makes them extremely efficient in space, where there's no gravity or air resistance, and slowly accelerating a spacecraft for weeks adds up the velocity, allowing it to travel at incredibly high velocities after just a couple of weeks of constant acceleration. A chemical rocket has a maximum velocity of around 40,000 mph. On the other hand, ion engines reach a velocity of 100,000mph, and NASA is researching and developing an ion engine which can theoretically reach up to 500,000mph.[3] Ion engines are already used in many satellites in order to keep them in orbit, and were used to propel NASA's Dawn spacecraft (launched in 2007) deep into our solar system. Dawn was the first spacecraft to orbit two objects in the asteroid belt between Mars and Jupiter: the protoplanets Vesta and Ceres.

To summarise, chemical engines are used up very quickly. The Falcon Heavy is about 550 metric tonnes, and 400 metric tonnes of it is fuel. This fuel is entirely consumed in 9.5 minutes[1], at which point the payload is in space. Whilst chemical rockets are still the easiest way to send objects into space, a typical mission to Mars will take about 9 months, because the spacecraft cannot afford to use any more fuel to accelerate in space, and will simply cruise its way to Mars. However, spacecraft utilising both chemical rockets to enter space, and ion engines to gain additional velocity in space, can reach Mars in just over a month, cutting travel times by nearly 90%. Ion engines hold lots of potential, and NASA are in the final stages of testing and improving their endurance and reliability, so we can use them on manned spacecraft.



We already know that even though chemical rockets are our only way of going to space, they are very expensive and relatively inefficient. Luckily for us, there could be another way to go to space: Skyhooks. Skyhooks are a simple and extremely effective method of sending objects into space. Currently, a rocket has to travel about 400km to deliver a payload into low orbit, such as the ISS. This takes a considerable amount of fuel and money, even more so if you are trying to fly to another planet. A skyhook could make it so you only have to fly 20% of the distance, and the skyhook would throw you into space. But what actually is a skyhook?



Put simply, a skyhook is a giant spinning tether, with a counterweight on it. This superstructure would have two ends, the lowest about 80km above earth's surface, and the highest about 1000km above earth. The craft would dock to its tether at the lowest point, and gain angular momentum as the skyhook rotated, as well as orbiting earth. Once the tether has reached its highest point, the craft would be released and propelled into space, carrying extra velocity from the skyhook. This not only reduces the fuel needed to send satellites and such into space, but also allows us to travel quicker and further, as almost all fuel is used to get out of Earth's gravity well. If we can use this alternative, we can conserve our fuel and use it to travel in space itself faster. We can even build skyhooks to orbit the Moon, or even Mars, using one of its Moons as a counterweight, and making a much larger skyhook to catapult us to the far depths of our solar system, at a much more affordable price and shorter time frame.

Despite this seemingly archaic idea to catapult things into space, the science behind the skyhook is sound, and it is even possible to create it with our current technology! There are many materials that are contenders to be used in the making of the tether, with a material called Zylon[4] in the lead. Zylon is extremely heat-resistant, a crucial feature as the bottom of the space tether is still in the Earth's atmosphere, and will experience some air resistance and a lot of friction. Fortunately, at 80km above the Earth's surface, the atmosphere is very thin and air resistance is bearable for the superstructure. Zylon is also over one and half times stronger than Kevlar, and barely changes length when put under intense tensions. All very useful properties to face space debris and endurance testing. Using the law of conservation of momentum, every time a skyhook sends a spacecraft into space, in order for the spacecraft to move faster, the superstructure will impart some of its own momentum on the spacecraft, slowing its own momentum. If this continues to happen, the skyhook will stop rotating and may even fall out of the sky and crash down to earth. This is why there will be small boosters (similar to the ones currently on satellites) to replenish the skyhook's momentum. Another way we can keep the skyhooks in the air is by doing the reverse of what we already do, as this would give energy to the skyhook, essentially replenishing its angular momentum. This would occur by "catching" high speed spacecraft in high orbit as they come back to Earth, slowing them down, and releasing them in low orbit. If done right, this can make skyhooks entirely self-sustaining, meaning they become more cost-effective the more we use them.

Since skyhooks can reduce the flight distance by about 80%, rockets can also be around 80%[5] smaller and considerably cheaper, although they will be remodelled so they can fly to and get caught by skyhooks. With the addition of skyhooks on other celestial bodies too, skyhooks could make future space travel cheaper and even commercialised, with planets such as Mars only one skyhook away. Skyhooks on a much smaller scale have already been tested and put into space, with great success, so

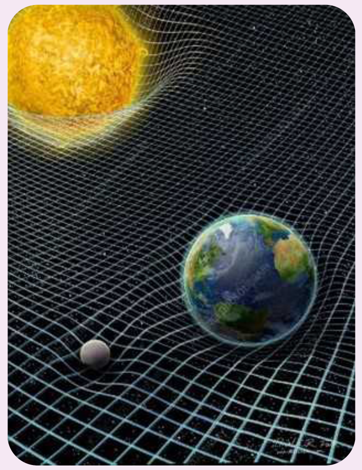
it's only a matter of time before space agencies invest in creating these superstructures. With a bit of luck, and a lot of hard work, we may see the commercialisation of space travel happen very soon.

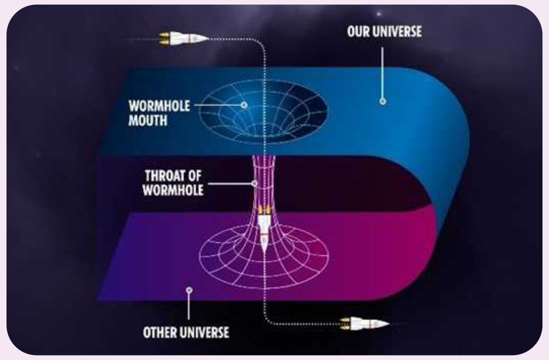
So far, we've seen technology that can be implemented within this century, with the concepts being very realistic. What about something more... theoretical?

Introducing wormholes.

Wormholes are a special phenomenon, which, according to Einstein's theory of general relativity, could exist. They are essentially doorways to a different part of the universe, whether that be a few metres apart, or billions of light years away. Using Einstein's theory of general relativity[6], we can see that the mathematics checks out, and that they can indeed exist. But just because something can exist in theory, it doesn't mean it necessarily does exist in reality, and we have not (yet) found any naturally existing wormholes. But first, some explanation and context of what wormholes and general relativity are.

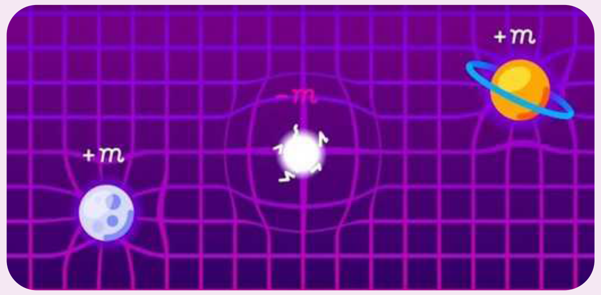
The theory of general relativity, simply put, is a way of describing how gravity works, and is our best description of it till this day. One key concept is how the very fabric of space-time works. Imagine if space-time is a trampoline. On its own, its flat, and there is no curvature. If I put a kettlebell on it, it will bend, and depending on the weight and shape of the kettlebell, it will curve differently. This is exactly how space-time works; objects such as stars and planets have a volume and gravity, and this gravity curves space-time. We already know that blackholes exist, and that they have such a strong gravity, warping space-time to the point where even light, the fastest thing in the universe, cannot escape its grasp.





But what if we took the trampoline and curved it so much that two opposite ends meet, and then we punch a hole through that? This is what wormholes do. They warp space-time so much that the space-time between two far apart places creases in on itself, and upon creating a hole through this crease, you have essentially closed the distance to virtually nothing. Another way of visualising this is if you think of the universe in 2D, as a big flat sheet. If this sheet is bent in just the right way, a wormhole can be created through the sheet. This method is technically faster than the speed of light itself, as you have taken a much shorter route.

As you can see, even though the maths behind wormholes checks out, in reality it does sound like science fiction. But what if instead of looking for wormholes, we try something crazier. We could try to make them ourselves.



In order to create a traversable wormhole (a wormhole we can freely travel through back and forth), we need three main criteria. Firstly, we need them to connect two distant places, such as Earth and Saturn. Secondly, they should have an entry and exist, without any event horizons, like blackholes, as that would make travel impossible. Lastly, they need to be big enough so that humans can travel through them without the G-forces killing us. The hardest part will be to keep our wormholes open. No matter how big our wormholes may be, the gravitational forces acting on the wormhole will be so strong that it will always be forced closed, unless we have some material that is strong enough to go against gravity to keep it open. This material has to have unique properties; the properties of exotic matter. Exotic matter is nothing like any matter we have seen previously, or even antimatter, which consists of the antiparticles of the particles of mass. Exotic matter has negative mass, meaning it will be able to repel any matter, and essentially repel against gravity itself. Surprisingly, this doesn't introduce any violation of the laws of conservation and energy, and are still mathematically consistent. This would be able to hold up our wormhole, and make not only space travel, but any sort of travel at all, reach new and unprecedented heights. There are even some exotic atoms and exotic hadrons, as well as other elements which may possess the properties of exotic matter.[7] However, some scientists believe that if you can travel faster than the speed of light, that means time travel should be possible, and this would break our fundamental laws of physics. For this reason, many believe that not wormholes cannot be found or created.

For now, wormholes exist only on paper, and in our dreams. But the future is a mystery, who knows what we could achieve later on... only time can tell.

The future of space travel knows no bounds, and an exciting world awaits us. From the ongoing projects of ion-engines to the dream of creating wormholes, who knows what we can do. Perhaps ion engines will make travelling to Mars as easy as catching a train, or skyhooks will let us travel to the Kuiper belt, where we can mine alluring new minerals, and even find exotic matter needed to open wormholes. The sky truly is the limit... or is it?

Asad
Ahmad

Is there hope for humanity to expand its authority onto other star systems?

No matter who you are, it is a universal dream to explore and experience new planets, exotic atmospheres, and alien scenes but is that really something we can do? The rapid pace at which technology is evolving is bringing forth the possibility of daily space travel, which earlier seemed to be a distant fantasy. This rate of development, with the aid of continuous innovation from private firms within the industry, raises the question of how long will we have to wait before star systems become within our grasp?

The viability of eventually traversing exoplanets, planets outside our solar system, relies on two main factors – our ability to detect them and proficiency in discovering the climate on these planets to test for their habitability. The second of these factors is made much easier once the first is tackled.

The easiest way to detect an exoplanet would just be to observe it. However, this is much more difficult than it first seems since we are able to perceive objects due to light reflecting off them. In the case of exoplanets, regardless of whether an atmosphere is present that reflects an abundance of light, the light that is reflected by the exoplanet will be overwhelmed by the light being emitted by the host star. It is analogous to holding a candle in front of stadium lights; although the flame is there, you cannot see it as it is inundated by the luminosity of the floodlights. As a result, we must explore three other methods for detecting and classifying exoplanets: Radial velocity, transits, and gravitational microlensing.

The radial velocity method is one of the simplest and relies on the interaction of gravitational fields, irrespective of the masses present. You may recall that every object with a mass is accompanied with a gravitational field and the heavier the object, the stronger the gravitational field. Therefore, a planet also has a gravitational effect on its star despite being a miniscule one. This causes a wobbling of the star and can in turn be used to detect exoplanets. Due to this wobble being more pronounced the closer a planet is and the larger the mass of the planet.

This method is most easily used on millisecond pulsars which are a special type of star formed from the remnants of a supernova and very dense and very radioactive so even if there are planets here, they are never going to be habitable. What is special about millisecond pulsars is that they release exact periodic pulses of radio waves which can be measured from Earth and plotted against time. If there exists a planet in orbit around a millisecond pulsar then it causes the pulses, every so often, to be offbeat.

This 'offbeat' pulse is regularly repeated thus giving astronomers the value of the period of the orbit the planet has around the pulsar. Using this and Formula 1, you can calculate the radius of the orbit of the planet.

$$r = \sqrt[3]{\frac{GM_*P^2}{4\pi^2}}$$

Formula 1

'r' is the radius of orbit; 'G' is Newton's gravitational constant; 'M' is the mass of the host star; 'P' is the time period of the orbit*

If the mass of the planet is also known, M_p then we can calculate the velocity of the star, v_* (Formula 2), in addition to the radius of orbit of the star, r_* (Formula 3), around the shared centre of the mass of the

$$v_* = \frac{2\pi r}{P}\frac{M_p}{M_*}$$

Formula 2

$$\frac{M_p}{M_*} = \frac{r_*}{r}$$

Formula 3

system.

However, there are some drawbacks to this method; most notably, it must be done on millisecond pulsars otherwise the observation time of the pulses will need to be over years rather than days or weeks. Secondly, if we are trying to find habitable planets then this is completely the wrong place to look as no life would survive under the harsh radioactive conditions. Thirdly, even if a planet is very close in, it produces a very small wobble so any planet further out, maybe in the goldilocks zone (a region of orbit around a star where it is possible for liquid water to exist), will not be detectable with this method as the wobble just would not be distinguishable. Fourthly, the planets found using this method are almost always hot Jupiters as they are both close in and are 'massive'. A hot Jupiter is a Jupiter-sized planet, which could be rocky or gaseous that is very close to its host star hence it is hot. Finally, using this method we have no way to prove that what we are measuring the effects of is a planet or if it is a dwarf star in a binary system with the pulsar. To prove that it is in fact a planet we need to know both its mass and density, we have mass but to find density we must use the next method, transit.

Transits are events that happen when an exoplanet passes in front of its host star relative to us. This leads to a slight dimming in the light that reaches us. Once per orbit, this dip can be measured to find the radius of the planet, r_p, where B is the normal brightness of the star, ΔB is the change in brightness and r* is the radius of the star.

$$r_p = r_* \sqrt{\frac{\Delta B}{B}}$$

Formula 4

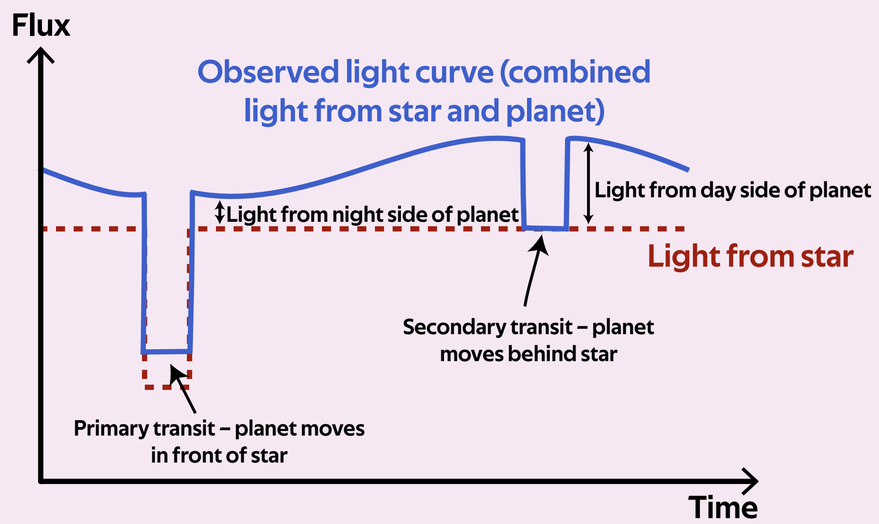
Using formula 1 we can also calculate the radius of orbit of the planet. Most importantly, we can find the density of the planet using both its mass and radius and verify that it is an exoplanet and not a dwarf star.

We can take this technique further to work out what the planet is made of as well, firstly by calculating its temperature, T, using the following equation ('L' is luminosity, 'D' is the diameter of the planet and 'σ' is the Stephan-Boltzmann constant in Formula 5) and also performing a spectroscopy on the light reflected by the planet; you can match the emission patterns to known chemicals and materials to gain a deeper insight into perhaps what it would look like on the surface.

$$T = \sqrt[4]{\frac{L}{16\pi\sigma D^2}}$$

Formula 5

Due to the following reasons, the spectroscopy must be done at specific times during the planet's orbit. For every transit there exists a secondary transit, which happens when the planet is behind the star and a secondary dip in brightness is observed. This is due to the fact that a planet is reflective so at every point except when the planet is behind the star it is also contributing very slightly to the light that reaches us. A graph of flux (a measure of brightness) against time is shown below:



Secondary transits are very difficult to do as the change in brightness is miniscule and indistinguishable from atmospheric disturbance in telescopes on Earth. Which is why space telescopes such as Hubble and Spitzer are used, recently the James Webb was also added to the fleet.

By subtracting a spectrum taken during the secondary transit from one taken just out of the secondary transit, the spectrum of the day side of the planet can be obtained and thus planets like J1719-1438 can be discovered. This planet also dubbed as the 'diamond planet' has around the mass of Jupiter but has the size of a core of a dwarf star, this extreme density causes massive pressures on the mostly carbon surface turning it into diamonds. Even though many planets have been discovered by Kepler, Spritzer, Hubble and many more using this method, they all have one flaw that is shared with the planets discovered using radial velocity measurements; almost all of them are hot Jupiters and if not then they are just extremely hot rocky planets. To help solve this and allow us to 'see' more Earth-like planets, we need to employ the phenomenon of gravitational microlensing.

Gravity, although the weakest of the four fundamental forces, can bend light by bending the fabric of space-time. In this way, when we look at a very distant object such as the bulge at the centre of our galaxy, and a massive object comes in between (for the sake of this explanation let's say it comes directly halfway in between) our line of sight, it distorts the image we perceive. This very effect is called gravitational microlensing and has been demonstrated in the example below.

Diagram 1 shows a singular ray of light from the focus (the bulge at the centre of the galaxy) as it is bent around the lens star leading to an increase in brightness of the image. If we instead look at every ray rather than singular ones, we see something like Image 1:

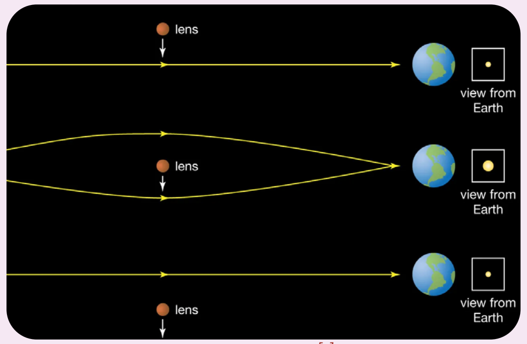


Diagram 1[1]

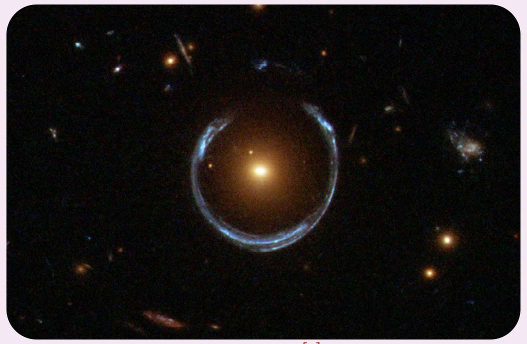
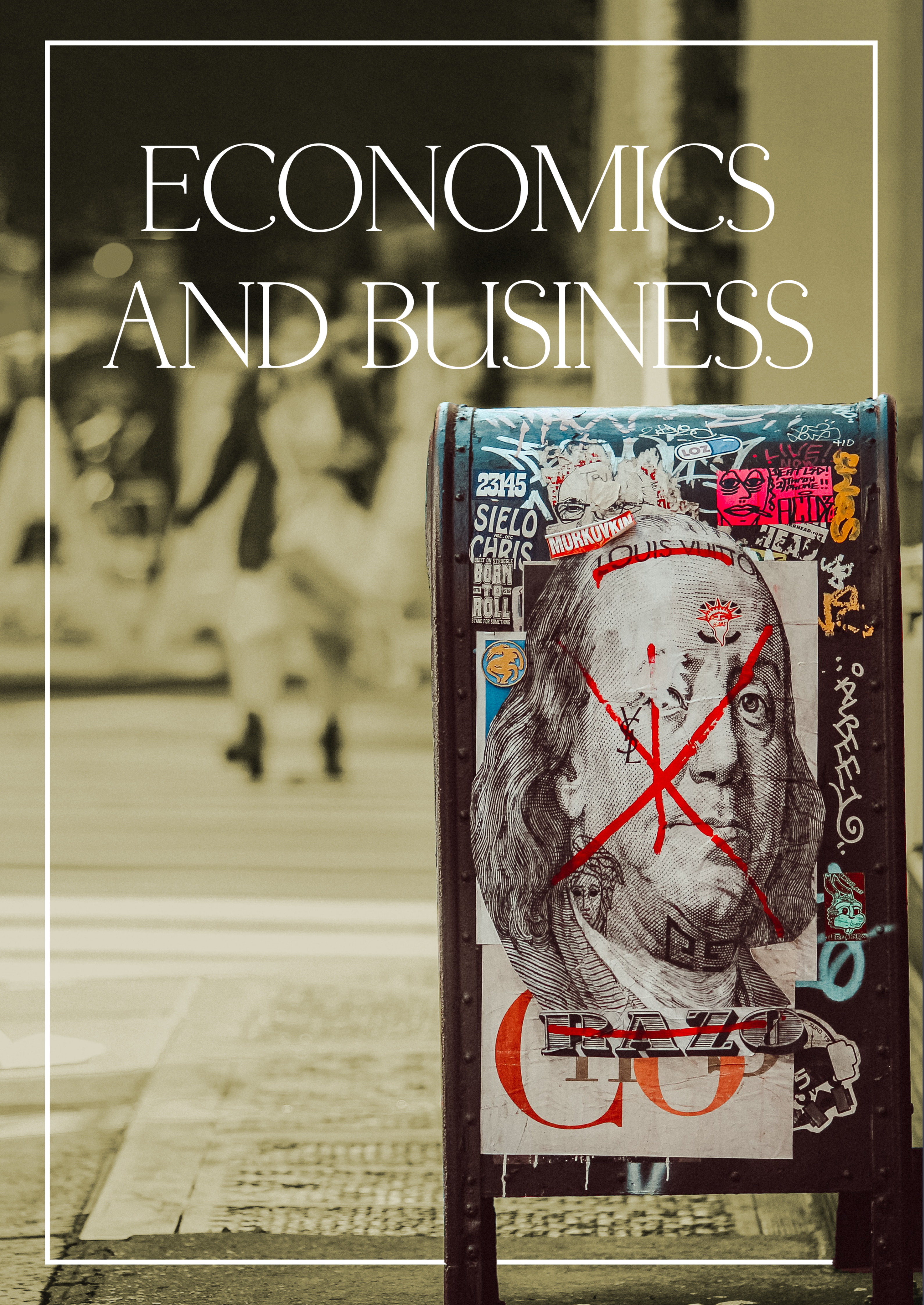


Image 1[2]

Since we know what the focus image behind the lens looks like 'normally', we are able to reverse engineer a lens shape to fit the scenario. The lens is what we are interested in as there are distinct patterns that emerge if the lens is just a lone star versus if it has a planet in its orbit. This is a very powerful, although rare, tool because we now have a way to detect the previously elusive 'non-Hot Jupiters' also known as Super-Earths or Mini-Neptunes which lie further out from the star. What seems to be microlensing's Achilles heel is the fact that it is a very rare event so while it may be our ace method, it is unreliable.

In conclusion, I think the idea of human beings leaving our solar system is currently absurd and will be for many hundreds of years mostly because, evident to this article, we wouldn't know where to go. There are many theoretical models for space travel but all of them focus on the journey rather than the end point, if we cannot find perfectly habitable planets close to us then I think it is a waste of resources to experiment ways to get there. It is much more prudent for Mars and Venus to be our top priority. On the other hand, we as a species are very adept at accelerating scientific progress at the cost of our planet, so who is to say we won't breakthrough again to find Earth 2.0?

ECONOMICS AND BUSINESS



PS

Parmeet Singh

An introduction to behavioural economics

Microeconomics is the study of decision making, and the interaction of producers, consumers and the government in markets. Bits and pieces of economic theory have been mentioned by the Ancient Greeks but in the late 1700s, modern economics came into existence as its own subject, gaining lots of traction until it became one of the most important social sciences in use today.

Neoclassical economics is, broadly speaking, the theory that focuses on demand and supply; the most important determinants of the production, consumption and market price of goods and services. This was revolutionary, as in the 1900s, it changed many economists' views and influenced classical theory, where costs of production was the most important factor for the market price of goods and services, among other ideas. Nowadays, when we think of economics, our brain goes straight to supply and demand, demonstrating the impact that Neoclassical economics has had on our views. But it's not only just our views, but our real world actions; our implementation of policy and how we run our businesses. But the Neoclassical theory has been around for a relatively long time from the birth of economics, and makes many assumptions, some of which aren't completely in line with our world today. This creates some problems when trying to implement this theory; And that's where Behavioural economics comes in.

Neoclassical economists believe in economic rationality. In theory, consumers always maximise their utility, or firms aim to maximise their long term profit, but there's so much more to it. When it came to how and why consumers make decisions, Neoclassical economics came up with a few principles, which together make up rational choice.

One axiom of this theory is that consumers process the information of an economic decision objectively and are not subject to framing manipulation, having fixed preferences.

Another assumption is that consumer preferences are complete, which means that the consumer will have a preference, and when presented with two goods will be able to say which they like better, or will be completely indifferent. They will not be undecided or get confused on which one to buy, they will either pick one, the other, or neither, and not for the sake of it.

The third, and possibly most famous, idea of rational choice is the Substitution axiom. It argues that if a consumer has no preference between two goods, then the consumer would be indifferent between two lotteries with those two goods as their prizes. Now this axiom further states that if something identical is removed from those two goods, the consumers would still have the same preference between them; known as the cancellation principle.

Neoclassical economics has so many rules and principles regarding consumer choice and rationality, and after describing a few of them one may question how Behavioural economics ties into any of this. Well, Behavioural economics revolves around the fact that consumers are not rational, and how they can potentially be manipulated into making certain decisions - decisions that consumers think they are making all on their own.

As with most parts of economics, Behavioural economics can be traced back to Adam Smith in the 1700s, who knew that people are overconfident and often overestimate their abilities, which are concepts that can be found in Behavioural economics today. However, it was much later, in the 1980s, when this came into fruition. Amos Tversky and Daniel Kahneman were very active psychologists during this time, identifying key concepts of Behavioural economics, and setting the foundation for future economists to build upon. Most famously, they found that people are subject to framing manipulations, with a theory they called “Prospect theory”.

For example, most would choose to win £250 guaranteed rather than gamble on a 25% chance of winning £1000 and a 75% chance of winning nothing. However, if presented with the chance to lose £750 guaranteed or a 75% chance to lose £1000 and a 25% chance to lose nothing, most people will risk losing £1000, hoping to lose nothing. What this shows is that people are more likely to take on the risk when trying to avoid a loss rather than gain a profit, which contradicts utility theory.

In the 1980s, Richard Thaler began collaborating with Tversky and Kahneman, and further built upon this work. It was around this period of time where Behavioural economics is said to have properly been created, and Thaler was credited with being its founder, later earning him a Nobel prize in Economic Sciences in 2017. Thaler built the base of Behavioural economics, most famously writing a book in 2008 about the concept of a “nudge”. This concept of a nudge, a way to influence the behaviour of an individual, is now synonymous with Behavioural economics, and it’s all thanks to Thaler.

Looking at real life examples we can see why neoclassical theories may be wrong when it comes to rationality, and Behavioural economics can, to some extent, explain why people do what they do. But Behavioural economics can also be used to help solve market failure. Market failure is where the market doesn’t deliver an efficient allocation of resources; goods are being over/under produced. This is against neoclassical theory as it is stated that markets will reach allocative efficiency, through the power of the “invisible hand”, and when market failure occurs regulation is typically the answer. However, as with any economic decision, there are opportunity costs, for example for the regulation to be effective monitoring and policing is needed. Additionally, bans could lead to illegal activity and regulation could also reduce consumer choice. And this is where the concept of the “nudge” becomes very useful.

In Thalers’ book he sets up a scene of a school cafeteria, and Carolyn, the director for food services, is trying to influence the decisions that the children make, without using typical measures like banning specific foods or likewise. She found that by placing certain foods at eye level, or making separate queues for desserts and main meals, she could increase or decrease the consumption of many food items by up to 25%. These simple small changes can be used to reduce market failure to a certain extent, and with potentially smaller opportunity costs.

However, while this may be used to reduce market failure it can also be used to achieve the opposite. Firms use many ideas of Behavioural economics when making decisions. For example, the concept of the default choice is continually used. Research shows that humans rarely deviate from the default choice, the option that is selected if the consumer selects nothing. Even if people are aware of a change to the default settings or can choose between the default option and another options, many people choose to stick with the default. This is exploited by firms, as if they find any one policy or such gives a better outcome for them, they will simply make it the default.

Another example of how firms use loss aversion to sell is quite commonly found on online shopping sites. Loss aversion is the idea that people dislike losing more than they enjoy winning, meaning they would choose much lower risk options, even if it means drastically reducing the reward, and this can be seen in the Prospect theory example we covered earlier. Shopping sites use phrases like “There’s only one left”, “Limited stock” or “Selling out quickly”, creating a sense of urgency and also fear of missing out. Loss aversion can also apply to social norms, and people trying to not lose in the sense of fitting in. If items are appearing to sell out quickly the risks of not buying are seemingly more than the risk of buying, rushing people to the checkout.

Looking at Behavioural economics on a surface level, it may seem like the solution to all of neoclassicals problems in microeconomics, as it stems from the fact that people don’t always make choices in their best interests, which is entirely what neoclassical theory is based on. However, the rise of Behavioural economics, while having solved problems, has also created problems. It is as writer and philosopher Robert Pirsig says: “every answer leads to ten more questions”. With the discovery of nudges and choice architecture, how the environment of the decision can be used to influence the decision, we realise that all of our decisions, whether intended or not, were made with us being nudged in some direction. And

with that we realise that it is impossible to not influence peoples’ decisions, and that there is no such thing as a neutral layout, because it is still a layout.

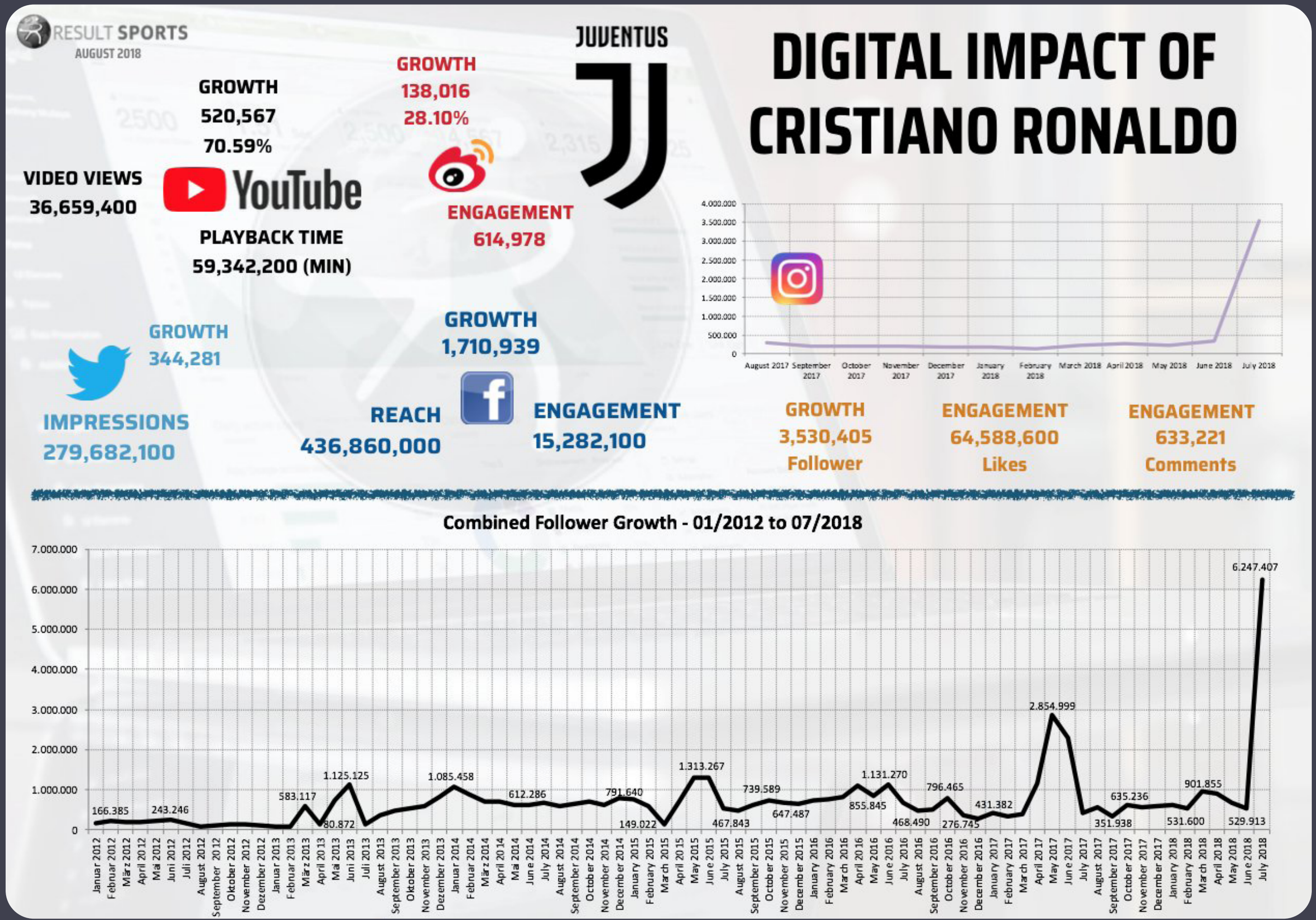
Finally, Behavioural economics is still a relatively new concept, and while 2008 may seem like a long time ago, there is still further to go until Behavioural economics will be able to make as much of an impact as neoclassical economics. Additionally, schools of thought such as Keynesian and neuro-economics have helped to make our understanding and practice of economics more informed. Our world is extremely complex, with there being so much that we don’t understand. Behavioural economics models should be used as a tool, alongside Neoclassical and Keynesian models, to help us make some sense of how our economy works, and what we should do to improve it.

Yuvraj Gautam

The economics behind football player prices

Since 2014, with all factors being equal, the annual inflation growth rate on the transfer market for the top 5 league footballers has been 26%. Reports also show that with respect to 2011, the same player costs almost three times more[1], but why?

The expensive nature of transfers originates from the marginal value of the very best football players being immense. Considering that football has risen to become the most popular sport in the world, the returns that these players can generate to their respective clubs is massive.

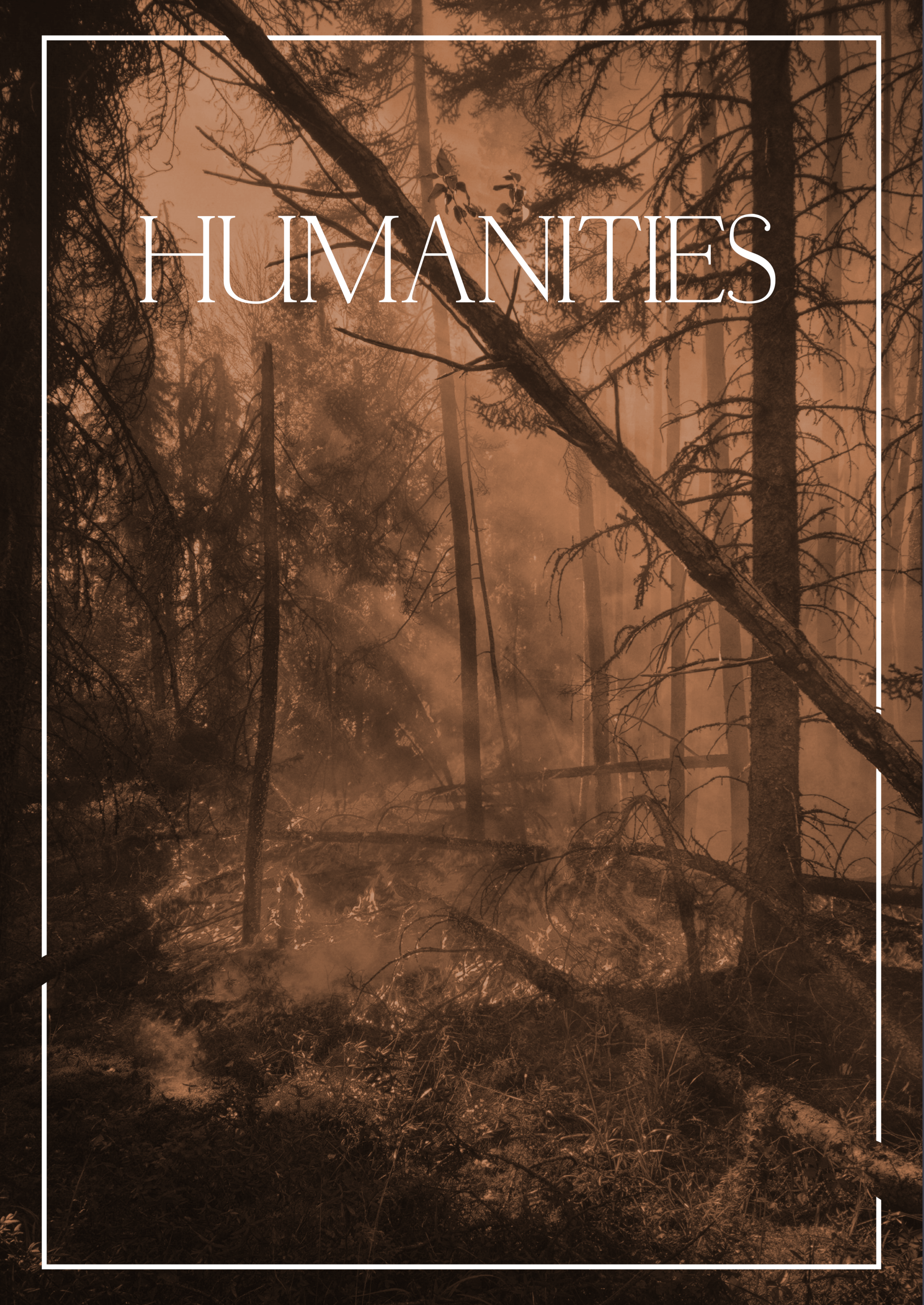


Looking from a football teams' perspective, there are 4 main revenue streams: sales from match days (tickets, food and associated merchandise), commercial sales (shirt sales) and revenue from TV licences. For most large clubs, match day sales are mostly fixed with little fluctuation since they sell out most to all their stadium. On the other hand, the other 3 factors are massively dependent on the success of the team and their overall popularity. Sponsorships tend to also be driven by the reputation and success, where more notable and prestigious teams land higher value sponsorships; Liverpool FC's deal with their defining sponsor Standard Chartered, a British bank, is rumoured to be worth around $59.2 million a year[2]. Commercial sales are a weird mix of popularity and success, for example Man United have recently been underperforming largely but have no doubt been constantly selling shirts and tickets due to strong fan loyalty. TV revenue is also determined by several ways differing by league/tournament, the Premier League for example determines the final allocation of TV money by league position, how many games were shown by the team and also by the pool of money generated from selling the rights domestically and overseas.

Overall, 3 of these income streams can be maximised by having the best possible team and players. Franchise players can significantly help considering that many of these players come from countries with a large domestic market or fanbase; in 2019 Juventus signed Ronaldo for a fee of approximately £105 million, but his shirt sales alone reportedly earnt around $60 million in just the first 24 hours[3] on top of the huge digital impact shown by the graph, proving how significant the ROI was associated with this transfer. It would be fair to expect the value of these franchise players to increase even further as more developing countries begin to participate more as global consumers of the games.

Another reason that accelerated increases in the value of transfers is due to a crazy transfer involving Neymar Jr, a Brazilian Prodigy. This all stems back to 2013, with Neymar joining Barcelona from Brazilian club Santos for a substantial but well reasonable fee of £49m4. Barcelona set his release clause at €222 million, which at the time seemed more than enough to put off any rational sporting investors, but PSG are not a rational club and in 2017 decided to break the transfer record by signing Neymar for this absurd fee. This sent the market for high tier football players into a vigorous spiral, with the first domino effects being seen in Spain where €1bn release clauses in renewed contracts have started to become more common with the top 4 teams. Desperate to replace Neymar with this huge sum of cash, Barcelona's first mistake was to purchase Ousmane Dembele for €105 million from Dortmund where he was injury plagued and only scored 19 goals in 102 games. Disappointed with his performance, and also needing to replace Iniesta, Barcelona spent a further £120 million on Coutinho from Liverpool who performed even worse and ended up being loaned out. Within 6 months, all the money that Barcelona had received was gone, both players signed did not live up to expectations, causing it to anger fans and lead to substantially less revenue being generated compared to what was expected, whilst causing every other decent attacker's value to be inflated by a huge margin, such as Joao Felix for £114 million and Jack Grealish for £110 million. There was also a surge in the number of free transfers; as more and more players were denied transfers to new clubs due to ridiculous valuations, many just started running down the time on their contracts and leaving like Pogba with Manchester United, Rudiger and Alaba with Real Madrid and Messi with PSG.

In general, it has been much more common to see higher than ever price tags being slapped onto even the most mediocre of talents, but this is not surprising in a highly competitive industry with highly profit maximising clubs that will take every opportunity to squeeze money out of their rivals and fans. With more clubs seeing the potential of franchise players to help their finances and the fall of an old generation of players, there is even more demand for new football superstars for young fans to idolise, support and financially contribute towards as they grow up.



Wildfires, and how they are affected by climate change

Jack Fraser


Wennington, West London, after a grassfire in July 2021

During the recent heatwave in Europe, fires broke out in various countries, such as the UK - where London saw "the busiest day for the fire service since WWII" - and France - where Bordeaux suffered multiple "monster" wildfires two months in a row, blazing through over 14,000 hectares in June and 7,000+ hectares in July.

Europe suffered through an event called a "heat dome," which occurs when high pressure sits over an area for an extended period of time, trapping a warm air mass beneath it. High pressure pushes air to the surface, as it prevents it from rising, creating increased temperatures through compression. When sufficient levels of each have been reached, a fire starts. Unfortunately, a heat dome also causes drought, which dries up the grass and provides more than enough fuel for a fire to start. Finally, oxygen is needed to stoke the fire and cause it to grow. These three elements - heat, oxygen and fuel - form the fire triangle.

Something still needs to ignite the fire, however. Natural causes, such as lightning, sparks from rockfall, volcanic eruptions and spontaneous combustion, can all start wildfires. Wildfires have been natural occurrences for approximately 419 million years, and their presence has had evolutionary effects on most ecosystems' plants and animals. Humans have also got ways of igniting fires, from sparks off machinery, power lines and trains, as well as people deliberately starting fires, either through arson, mistakes (like a smouldering cigarette or barbeque) or slash-and-burn (where a section of forest is razed to make way for construction). Preventing all these causes is futile, as there are far too many occurrences worldwide to stop them all, especially as a result of global warming, which makes the circumstances for fires more common, and increasing the amount of fires worldwide.

Perhaps the most interesting cause of ignition is spontaneous combustion. A flashpoint is when all three points of the fire triangle are present, and something ignites the fire. Flashpoints can be reached in other high-pressure, low water level areas too, such as compost bins, turning what was once innocuous into a ticking time bomb, should the temperature ever increase to the point where the compost bin self-ignites. In Canada, there is a common type of tree - the Ponderosa pine. The high sap content is an evolutionary advantage of the Ponderosa, as it allowed the tree to survive the fire season. However, global warming is increasing the temperature and causing the sap to evaporate. Without the liquid in the sap, the wood reaches a flashpoint and self-ignites.

When visiting Canada, on two occasions I was subject to some wildfires, likely caused by the Ponderosa pine self-igniting. Recently I drove through the small town of Lytton, which had been caught in the path of a fire and had burnt to the ground in 2021 - 90% of the town was destroyed, and over 2,000 people were affected. In Canada, most houses are constructed out of wood, due to their significant lumber industry, and Lytton was no different, facilitating the high destruction rate. The government had put barriers up to prevent onlookers, but through cracks I could see brown grass on black soil, and the outlines of roads and houses. Seeing the aftereffects of a wildfire was quite shocking, as most of the 6 hour drive north was accompanied by charred trunks and brown grass. At one point, I thought a mountain had snow on the top, which I thought was unlikely in July. I was corrected - the fires had burnt all the vegetation away, exposing the raw stone underneath.

The second time was significantly scarier. In 2017, Canada had one of its worst wildfire seasons in recent years, and I was right in the middle of one of these wildfires. Imagine looking out of the windscreen and seeing only 30 metres in front of you. With a few cats eyes on the road to guide us through the wall of smoke that reached so high the sky was tinged red due to all the light from the fire, it was difficult to see where to go, and our speed was reduced to a fraction of what it should be on a freeway. It was quite an experience, and the smoke from the fire reached all the way south to Vancouver, the largest city in British Columbia.

Now imagine how dangerous an undetectable fire could be. Detection methods on a large scale or in a rural area consist of firewatch towers and satellite images, which, naturally, cannot detect underground fires. Coal seam fires and root fires both travel, undetected, underground, for up to hundreds of miles. Fires may be put out on the surface, but can reappear days later when a root fire travels upward through a tree trunk or a coal seam fire comes too close to the surface. These kinds of fires are rare, but are inherently dangerous, as the short time to react means they can cause extreme damage, both short-term and long-term.

Starting by taking a look at the short-term effects, wildfires suck cold air down as hot air rises, due to the pressure differences created, which feeds the fire with oxygen and ensures it keeps going. This movement of air also creates wind, which can blow the fire at speeds of up to 6.7mph in dense forests and 14mph in grasslands, a danger for anything in the path, and sometimes outrunning evacuation efforts, as in Lytton, where sadly two inhabitants of the town died. The damages can prove costly, especially for weaker economies that cannot deal with the long-term rebuilding costs. The Lytton fire has cost US$150 million (as of October 2021), and the town is still not rebuilt, meaning the 2,000+ people affected are still displaced.


São Paulo, cloaked in smoke in 2019

Smoke is also an issue, both short and long-term. Short-term, it can lower visibility and prevent effective evacuation and medical assistance. In Canada, where distances are vast and terrain is vertical due to the Rocky Mountains, helicopters are used to rescue people trapped. However, smoke obscures vision and makes flight incredibly difficult. Smoke can be carried by the wind and blown many miles away, such as the August 2019 Amazon wildfires in Bolivia, where the wind carried smoke over 900 miles to São Paulo in Brazil, engulfing the city in darkness in the middle of the day. The smoke also brought much lower air quality, affecting those with breathing problems especially badly, and lowered visibility to the point where basic travel became much more difficult to achieve, preventing people from going to work or visiting those affected by the smoke. If allowed to enter a building or other enclosed space, the smoke can also cause carbon monoxide poisoning too.

Smoke will clear after a short period of time, and houses can be rebuilt. Long-term effects are much trickier to deal with. Wildfires are a negative feedback loop - this means the effects from wildfires cause an increase in wildfires. Trees contain absorbed carbon from the atmosphere, and act as carbon sinks. They retain this carbon for years until they die, and a trees' age is usually in the triple-digit range. When trees die, they release some of this carbon into the ground as it decomposes, and a small amount returns back to the atmosphere. When a tree burns in a wildfire, the carbon isn't lost through decomposition. Instead, almost all of the absorbed carbon returns to the atmosphere. This helps the enhanced greenhouse effect, where gases like carbon dioxide trap heat from the Sun by scattering the rays and preventing some of the rays from reflecting off the Earth and going back out into space, increasing the temperature of the planet, and increasing the risk of wildfires.

WIldfires, naturally, cause environmental damage, especially when multiple bad wildfire seasons follow one another, leaving no time for the environment to recover. Given time, a habitat can spring back in just a few years, as plants have adapted to live with wildfires. Sequoia trees, for example, rely on fires to reduce competition for water, nutrients and sunlight, to release seeds from their cones and to clear the ground, ready for new growth. New growth cannot happen, however, if the frequency of wildfires increases, preventing those species adapted to live with wildfires to deal with the increase in intensity and frequency caused by climate change. If plants that have adapted to live with fires can't

cope, the effect in areas with plants that haven't adapted to deal with fires will leave longer-lasting damage. Climate change also increases the reach of wildfires, away from the equator and tropics, increasing the amount of regions that will experience fires. Most of Europe falls into this category.

So, how will Europe cope with this new threat? There are many ways to detect wildfires, and many ways to put them out, but few ways to prevent them.One of the few ways is to warn or ban ignition culprits like barbeques and smoking in wooded or grassy areas, as well as educating users of these spaces to the dangers of left-over smouldering litter. Detection methods are going to need to be adopted too. Large-scale methods like firewatch towers or satellite imagery are relatively simple to implement, but limited in their ability to detect smaller fires. They are also reactionary, meaning the fire will have had time to grow and spread before it is detected, which gives the fire a chance to grow larger or do some damage before it is noticed. Smaller-scale detection methods include remote sensors, which will be useful in high-value locations, such as population centres or wildlife preserves. This method can detect fires before they are given a chance to grow large and deal significant damage, but is costly to implement, limiting its large-scale effectiveness. Finally, weather data can be used to try and predict when wildfires have the highest chance of occurring, allowing wildfire detection methods to be put on high alert. However, as the recent heatwave has shown, changing weather over short periods of time can be exacerbated by climate change, changing rapidly from year to year, meaning estimates may not be as accurate as they would be otherwise and limiting the effectiveness of using weather data to make predictions. Additionally, due to this rapid changing, archival data is less accurate and prevents the accuracy of these predictions, as predictions are drawn from a smaller data set.

As wildfires prove inevitable in the face of difficult prevention methods and limited detection methods, effective methods of extinguishing the flames will be needed in the future. A common sight in wildfires, especially in countries with large areas of affected land, is the water bomber. These planes scoop up water from lakes and rivers, and drop them over the fires to extinguish them. Effective in rural locations and for large fires, but difficult to use if no water source is nearby and can create havoc if thousands of litres of water is dropped over an urban area. Their effectiveness in rural areas is due to their ability to hold up to 74,000 litres of water, although most bombers carry much less than the 747 Supertanker, which set the record for largest rated capacity in a water bomber. The Russian Beriyev Be-200, pictured below fighting fires in Israel in 2010, can carry 12,000 litres.


The Russian Beriev Be-200 heavy aircraft, fighting fires in Israel in 2010

Other prevention methods exist, such as natural or man-made barriers. These generally consist of lakes, wide rivers, roads (generally motorways, although it can be smaller roads too) and areas of little vegetation. These barriers prove effective if wide enough and if the fire is small enough, however overhanging fauna or large fires can cross these barriers, making them only marginally effective. The most effective prevention method, especially in urban areas, are firefighters on the ground, who have been trained to deal with large-scale urban fires or wildfires. Combined efforts from the water bombers to extinguish large areas and firefighters on the ground to attack the smaller flames the plane did not catch are the best ways to deal with these large fires.

To recap, wildfires are increasing in range, frequency and are becoming more extreme. At the same time, we are becoming more able to detect and prevent the spread of wildfires, through new technologies, such as new prediction models based on simulated weather forecasts in the future, allowing us to be even more accurate in predicting when and where wildfires are most likely to occur.

Ultimately, if any lesson is to be learnt here, it's that wildfires are not to be taken lightly, and will continue to be a threat now and in the future, and we must learn to adapt and mitigate their effects as they change over time. Recently, one example of this is the European firefighters working together in Bordeaux to fight the "monster" fires, sharing their knowledge as other countries that suffer from fires more frequently have deployed firefighters in France. This kind of coordination and teamwork is vital in adapting to cope with fires in areas where they have been absent until now, as it will become a more common occurrence as our world heats up due to climate change.

Antonio
'Hou Ting' Mak

Lessons from the 18th Chopin piano competition

The International Chopin Piano Competition is held every 5 years to honour the composer Fryderyk Chopin. It is a way for talented young pianists to start their careers through recording contracts and tours. Last year, 87 pianists from 18 different countries competed in 3 stages for a spot in the final round. All the candidates played exceptionally, but what could we take from them to improve our own playing?

I – Play the Correct Notes

Not really a revolutionary idea, but last year was a great reminder that the first step to playing at an elite level is to play all the correct notes whilst making little to no mistakes. No omission of the correct note or inclusion of a wrong note will make a high scoring performance.

Obviously, in other musical contexts, a wrong note can be worked around. To quote Miles Davis: "When you hit a wrong note, it's the next note that makes it good or bad".

However, in the context of 21st century piano competitions, to play a wrong note is to sin. It is the quickest way to notify the jurors that more time could have been spent working on the musical skeleton of the piece. Familiarising yourself with the pitches of the music is the highest priority before adding your own interpretational ideas to the piece.

Obviously, fruity things happen on stage: perhaps you can feel a member of the jury really laying down judgement on you for your hair, maybe your mum is in the crowd somewhere, maybe you just don't have a lot of experience. Whatever the reason, we can get nervous and make mistakes as a result. That's just one of those things that happens. Even at the highest level, those performers are still human; depending on the musical context and degree of error, the mistake may go unnoticed. When you make a mistake in a performance setting, the best response is to ignore it and recover the piece.

However, in a practice setting, we should really be aiming to be as perfect as possible. When you make a mistake during practice, the best response is to go back and correct it such that you minimise the chance of making the same mistake again. This also ensures that we learn the music as written – reducing our chance of producing errors on stage.

Practice slowly and diligently. Our art form is a discipline, first and foremost. To quote a certain concert pianist when asked why he never played wrong notes, "I don't practice them". The need for self control cannot be understated here, so going slow and steady with a metronome is ideal. If you are recapitulating lines of music without any attention to exactly what you hear, that's not really practising – that's playing.

Our goal when practising should be to play through something without making mistakes – this means we cannot approximate or guess anything: not the time, not the melody, not the fingering. Every detail should be deliberate and controlled.

The amount of music you end up playing doesn't matter. During practice, the only metric you should be counting is the number of mistakes you make. Graham Fitch emphasises the quality of practice being important: suggesting a routine of smaller, focused practice sessions of 20-30 minutes each. The bottom line is if you cannot play a piece through slowly with no mistakes, you cannot assume that you can do so at tempo, whilst adding artistic flair and interpretation in a performance setting.

II – Clarity as a Tool

Upon watching Bruce Liu's final round performance, the first thing someone might notice is how blazingly fast Liu can play. It becomes really obvious that this guy goes at whatever speed he wants because of his technical capacity. While his competitors might be surviving through certain passages, Liu exercises his talent-founded prerogative to play them however he likes. It's easy to forget or outright dismiss other merits of Liu's playing given his captivating dexterity.

Complementing Liu's demonic speed and fluidity is his crystalline clarity. We might talk about a melody being particularly embellished with cascading runs and exciting turns but Liu really does play the ornaments – there are no real embellishments. The "embellishments" are considered part of the melody.

Clarity is the result of a strong technical foundation combined with true legato. Because of that, it allows us to communicate that we are technically fluent and mechanically able to express ourselves on our instrument – it gives your performance a virtuosic spice. More importantly, clarity helps communicate melodic ideas to our audience. Musical ideas may not form as a result of clarity, but they are often greatly aided by it; if the audience cannot clearly make out what we are playing, we cannot hope to communicate our musical ideas to them.

Liu is clearly an incredibly talented individual and plays better than the majority of us could ever hope to play. In an interview for The Korea Times, Liu humbly stated that piano is just a "hobby" which he carries lots of passion for. However, is there anything that we could take away from his playing?

We often are easily able to achieve the connection part of legato by allowing our notes to overlap each other. However, that makes our playing hazy and we lose clarity. One way to improve clarity from practice is to practise our melodies staccato. After a few times, we can gradually lengthen the duration of the notes we play until they are connected without overlapping. At this point we will have achieved true legato with perfect clarity.

III – The Need for Musicality

Fans of Aimi Kobayashi may look back to her showing in 2015 and remember the agitation and fire she played with. A raw, emotionally turbulent player who dictated rubato, dynamics, and articulation to create a crude sensation.

In contrast, the Aimi Kobayashi who came out last year was much more controlled and deliberate. She still encapsulates reverie and effect extremely well, but there's a hint of refinement in it. The kind of control that you'd only expect from someone who could perceive time slower – that's what she showed us last year.

In her performance of the Concerto in E minor, op.11 in the final round, Kobayashi played at a much slower tempo than the accepted interpretation. She chose to throw away virtuosity in exchange for delicate, contoured playing. It's at this slower tempo that Kobayashi shows off her ability to craft and shape melodic lines into something beyond the standard often displayed at the highest level. Kobayashi's delicate touches dance effortlessly across the higher register and cascade melodiously downwards. Despite the slower tempo, it's still very much Kobayashi: emotional playing directed meticulously by the individual fingers; she's just being honest about her interpretation. It's a brave thing to go against the grain in any context, but especially so in a well-established piece at an international competition. From a competitive standpoint it, falls short – Kobayashi misses out on opportunities to show off her technicality; from a musical standpoint, her gambit paid dividends. Although it's not quite the technical spectacle of Bruce Liu, it's utterly beautiful.

Some identifying characteristics of Chopin's work are the lyrical melodies which often sit upon a strong

harmonic accompaniment which really carry his heightened level of emotional expression. Kobayashi takes this writing and invites herself to freely express herself musically across the piano.
I think what we could all learn from her performance is how the temperament of your nature can affect your playing. Sometimes it may seem musically inappropriate, but there is a reward to allowing yourself to actually play more. Of course, it is also important to recognise when you are overplaying, but that's something that comes with time, as it did for Kobayashi.

In practice we can work on our musicality by really playing to our ear's decree. "Is what I'm playing really what I want it to sound like?" If we find ourselves struggling, we can turn to other sources for inspiration: "how does famous performer A play this part? How does famous performer B do it? Which do I prefer?". Find out how your teacher might play that line. Questions like these can shape our interpretations immensely. Fuss over every detail to make sure your work is a true expression of self.

IV – Communication is Key

As Tiffany Poon noted in a 2020 interview with Bachtrack, classical music has a "disconnect between the audience and the music. The composers are dead and there just isn't that human connection anymore." She's profoundly right.

Hayato Sumino might not have made it into the final round, but he definitely reached a lot of people with his communicative instinct. His first round performance has so far garnered 2.8 million views, the most of any non-final round performance.

Listening to Sumino is really easy. He writes all his feelings and nuances down on a piece of paper for you to look at and read. Although Kobayashi is musically brilliant, it's not quite as near to us as Sumino. And I feel like that's exactly what makes him so great: Sumino has this ability to speak to us through his playing; every little articulation carries that spark of life which, in turn, gives his performances a refreshing vitality.

Sometimes, we have a specific narrative to convey if we are playing program music. Of course, in this scenario, our job as the musicians is to pass on the extramusical meaning to our audience.

Sometimes we are lucky enough to have a creative motive accessible to us; in this case it might be worth trying to come up with your own related narrative to communicate back to the audience.

If we are very unfortunate, we might not have the background or creative motive behind a piece accessible to us either. In such cases, a safe but liberating approach is to extrapolate whatever emotions naturally come to you from the music.

I think for a lot of us musicians, we don't really struggle with conveying any given emotion on command. It's just the ambiguity of what we are meant to do given source material which has gone through potentially hundreds of years. My interpretation cannot bring Schubert's very will and message back.

A first step towards improving your communication on your instrument is to define exactly what emotion you, the performer, are trying to convey – not the emotion that the composer wanted to convey, and certainly not whatever emotion the audience will extrapolate from you. We want to keep the creative vision in our mind transparent, not sentimentally self-indulgent, and most importantly – it should largely come from us.

Once we have defined our creative vision, we should then be able to use our sonar palette to shape our interpretation into a refined product. As clichéd as it sounds, there isn't really an incorrect answer here - but it has to be with the goal of communicating. "Does that tenuto actually help us convey an emotion or does it feel good to play?" The clearer your creative vision is, the easier this part should be.

It's important to understand that your interpretation can change.Allow yourself to change bits of it – allow your musicality to guide your creative choices, including how you choose to communicate a piece.

Closing Comments

As Adam Neely famously thrice said, "repetition legitimises". With that in spirit, let's quickly recap the 4 main points:

- Before refining an interpretation, practise slowly and make sure you are playing the correct notes
- Improve your clarity by practising the quality of legato
- Deeply consider your interpretation of the music, and let your musicality guide it
- Don't forget to make sure you do a good job of communicating your interpretation

The International Chopin Piano Competition is one of the world's most prestigious piano competitions, which are becoming increasingly competitive each year. The propagation of high fidelity, flawless recordings has led to concert musicians at competitions seeing their standards rise. This article is horrifically guilty of having indulged in music as a sport and a stark reminder is due to let us all remember that music is an art form before anything else.

I hope this article serves as a way for you to improve your own playing. If my suggestions don't work for you, I would at least hope that this article serves to allude to particular elements of your own playing - and that you seek to improve upon those particular elements.

Charles Philpott

Why your "bad" music taste probably isn't your fault

Music is a uniting force in a society where chasms grow ever deeper. It gives life rhythm. It's what the world relies on to keep on turning. But with nearly 8 billion humans on Earth, our tastes are infinitely unique. Join me as I shred sound to get to the heart of why we are who we are.

Introduction

To the minority of readers still with me after this statement, I assure you that you will be glad to have stayed. I wish not to insult, judge, marginalise, or criticise the one uniqueness we all have in this current period of time. In fact, it is the total opposite. I aim to open your eyes to the psychology and sociology behind music. The way that tiny vibrations in air impact our lives can seem like a web too complex to unravel, with connections in all areas. Religion, culture, media, sport; the list is endless. But with so much variation in genres and experiences, nobody is safe from the criticism of The Masses. This article will explore our relationship with music, why we like what we do, why society says it's wrong, and how, no matter how hard we try, we cannot change our musical taste.

The development of Musical You

Like all tastes, musical preference comes from a plethora of sources. It could be that you discovered disco from a diner in Dublin or revelled in rock in Reading. Our experiences can be as small as a snippet or as long as "Echoes" by Pink Floyd. A musical idea implants itself from the very second it enters our brain, which makes our younger years a melting pot of sound. Our control over what we hear is very limited at a young age, but when we grow older, we gain records, CD's, and Spotify to satiate our discographic desires. Therefore, our initial uncontrolled impressions can develop and grow as we gain the ability to listen to what we please. However, it is argued to an extent that our likes are partly predefined, linked to our very personality.

It has long been speculated by the intelligentsia that listening to Jazz was only for minds that understood it, hence the stereotype that with academic prowess comes a love of Coltrane. In a study by the University of Texas, they sought out to break down whether personality types affected our musical taste. Using categories of reflective and complex (jazz, classical, and folk), upbeat and conventional (country and pop), intense and rebellious (rock, alt, and punk) and energetic and rhythmic (hip-hop, soul, and funk), they found striking similarities between taste and personality down to details as small as political ideology. For example, if a participant listened to "intense and rebellious" music, a correlation shows they are supportive of the opposition political party to the one in power. Furthermore, those who listen to pop are extroverted and agreeable, but socially and emotionally unstable and those with liberal social ideas do in fact prefer Jazz. Therefore, your music taste is in part out of your hands entirely, with a significant portion being down to your personality.

If you are female between 11 and 14, or male between 13 and 16, go listen to more music now! Seriously, stop reading and go diversify your taste! According to evidence by Stan Stephens-Davidowitz's study for the New York Times, these are your crucial ages for the creation of your preferences. In studying every Billboard chart-topping track from 1960-2000 and matching it with Spotify data , the favourites of different ages become apparent. The classic teenage tropes of becoming independent and forging your own path are just as applicable to music. Consider the song Creep by Radiohead for example (I can already hear the complaints flying in). The song sits as the 164th most popular song amongst men who are currently 43 years old, but if we look at men 10 years older and younger, the song doesn't even clock in the top 300. Here, it is important to note that the song was released in 1993, when our Radiohead fans were 14 years old, directly in the middle of our age range for men. It is not just isolated to alt indie

90's rock either – this can be seen in simply talking to your parents about their music taste. The crucial aspect is not entirely when the song was released, rather when you are first introduced to that genre or artist. Just because your Mum's favourite song is Hound Dog, doesn't mean she is necessarily 75 years old , just that Elvis came into her musical radar in the above time period. However, the societal pressure of listening to that new album is what drives the previous, and most other examples. Now don't panic too much, there is still something you can do if you so wish to swap Swift for Slipknot. The study concluded that exposure in our early 20's is approximately half as influential than that in our teens, with the effect trailing off as one gets older. This continues on until we die, with very small adjustments being made the older we become.

By 33 years old, studies show that people stop seeking out new genres of music and slowly gravitate back to the tunes of their teenage years, seeking the youthful vagrancy, joy and excitement they experienced and therefore associate with the music. When pensioners are studied, this age they seek out is slightly later, in their early 20's, when their life starts to stabilise and become more carefree. In short, once we pass our crucial years of development, we use our previous music taste to try and return to those years we so dearly miss, instead of broadening our taste to experience something new.

Lies, lies and more lies

Our music taste is one of the things that comes up time and time again in conversational small talk, whether it be with a friend, colleague, boss, or stranger. It gives us an opportunity to express our likes and dislikes in a supposedly non-judgemental space where we are free to fight The Who against the Supremes. Due to this reflection that certain musical genres are associated with favourable personality types (and this is true to some extent), many people take it to the extreme. An international survey of 8000 adults exposed that 42% of adults will lie about the songs and artists they have listened to in order to seem more sophisticated or rebellious in what is essentially a futile scenario. We also give the excuse of "not wanting to hurt someone's feelings" when saying we have listened to an artist we have not, with 40% citing this as the reason. This deceit continues into relationships, where 20% of those who lied stating it was to impress a potential partner on a first date. But is lying really all that bad? Over two thirds of those who stated they had lied, then went on to listen to that song and this will inevitably open up a new possibility for one to affirm their taste, or perhaps discover something they never knew they had liked. This is the beauty of music. Due to our near constant exposure to new sounds, we can have a "musical gap" in our minds which can be filled later in life. One who listens to Jazz and Funk as a teenager will almost certainly enjoy Jamiroquai as an adult even if not experienced before. This is due to how interconnected styles can be and how the enjoyment of certain features are subconsciously stores as a part of your musical taste, awaiting discovery later in life.

But my taste is better!

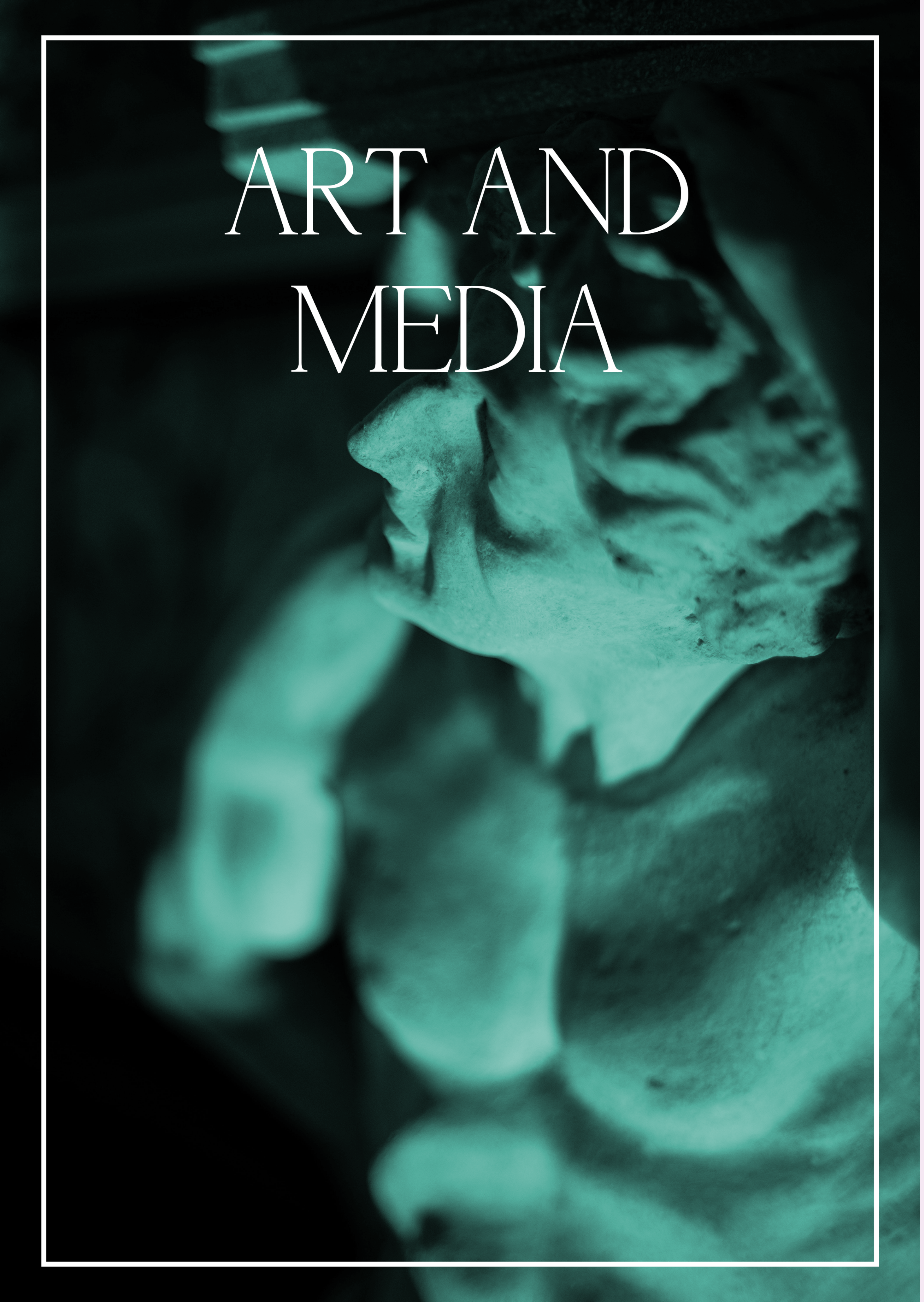
Every one of us has met that person who is just better than you because of their taste in music – and if you haven't, well, chances are you are that person. Regardless of your class, social standing, income, or appearance, some simply consider themselves superior because their little air waves in a specific order are better.

This is the biggest fallacy you will ever be told.

There is no objective way to qualify music as good or bad other than the number of people to enjoy it. I for one am not a fan of Cubano Pop due to my experiences, cultural upbringing, and exposure. But if I were to claim it was terrible, I would not be thinking rationally at all. There are millions of fans, and it is as futile as trying to say if a food is good or not. In fact, it is even more futile. Whilst many rightly argue that "without music, what are we?", without food, we are dead. Therefore, if we cannot agree on which foods are the best to keep us alive due to biological necessity and which are the most enjoyable, how can we even begin to comprehend which music "is best for our survival and culture"? Alas, musical purists continue to perpetuate stereotypes to justify their stance. So called "Tasteless" artists like Taylor Swift, 1D, or Billie Eilish are mainly enjoyed by teenage girls which means it is instantly devoid of value according to society. This ridiculing leaves people feeling isolated from what they enjoy, lest they upset the musical Big Brother, and it can ruin the experience of musical at the most critical time in somebody's life.

The reason behind this obtuseness is simply the passage of time. As time goes on, we remember almost all of the absolute classics, the chart toppers, the "bangers", and a couple of "tragic" songs mocked by the majority. Even these tragedies become popular as the reputation spreads and people listen to it more. Nowadays, the "bad" coexists with the "good" and in 20 years' time the cycle will repeat itself, but for now we the individuals are stuck with the criticism of the Musical Elite and sometimes the general public as well. Therefore, since your musical ideals will always be held as "bad" for so many, criticism is inescapable, but the joy lies in realising the depth of musical taste, its links to our personality and how it develops in the first place.

Then, to summarise, whilst your musical taste will never be identical to anyone else's, sparking criticism, and it is almost entirely set in stone, we all still relish and savour music. So, ignore the criticism, broaden your taste whilst you still somewhat can, and re-engage with your musical humanity. Don't lie about your taste too, but if you do, make good out of it. In attempting to understand and engage with what others find so dear, we better our musical and in fact our entire selves. In essence "All I'm asking, is for a little respect."



ART AND MEDIA

Shruthi Lakshmanan

Ever heard of 2.5D?

Animating in 2.5D (2.5 dimensional) is a fairly common technique in the industry - with it being used in entertainment ranging from our beloved Disney classics to video games. We know that 2D is a visual technique in which the images, characters and sprites are flat, whereas 3D computer graphics (aka CGI - computer generated images) make the digital object seem as though it is moving in three-dimensional space. So what is 2.5D?

It is, as the name would suggest, a combination of the two techniques - animators would mix 2D elements with 3D ones such as a 3D environment setting + a 2D character moving through it or even a 3D simulation that is then ultimately hand-drawn and made 2D.

2.5D Beauty and the Beast

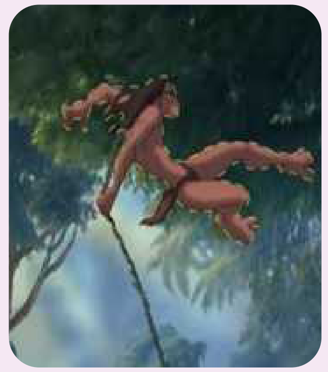
3D Raya and the Last Dragon

2D Peter Pan

2.5D Paper Mario

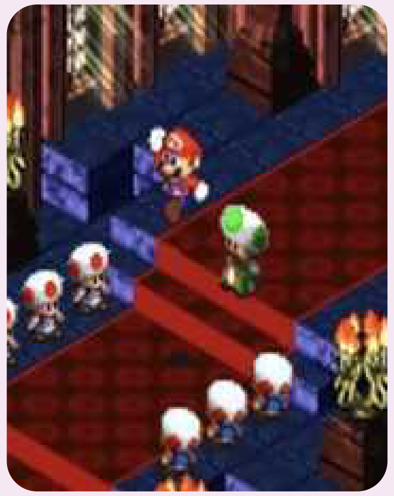
There has been a lot of controversial development in the animation industry over the years - the rise of 3D animation being seen as a replacement for nostalgic 2D films like Bambi (1942) - however 2.5D films pose a solution to the never-ending debate. This concoction seems to possess the fantastic details and lighting from CGI with the seal of a traditional hand-drawn feel. There are certain aspects that computers simply can't replace such as the 'comic strip'-like sensation depicted on the left.

Mixing animation styles is not a new concept and is quite prevalent in successful, modern day films such as Sony's Spider-Man: Into the Spider-Verse (2018) and Disney's Chip 'n Dale: Rescue Rangers (2022), as well as believe-it-or-not The Lion King (1994) and Tarzan (1999). Probing further into the last two, 3D techniques were often used to simulate a large movement or environment, e.g. the wildebeest stampede scene in The Lion King, whilst the entire Tarzan jungle was created in 3D with a hand drawn Tarzan swinging through. The pioneering computer animation software they used at the time for this effect was Deep Canvas.



In fact, the Oscar-nominated Apple TV film "Wolf-walkers" (2020) used 3D models and VR to create a CGI forest much like Tarzan. However, the animators took this one step further and recreated each scene by hand, using graphite on thousands of sheets of paper - to mimic art from the 1600s. Ultimately, 2.5D serves as an in-between medium to make use of the best of both worlds. "The idea was that it would not feel too clinical, too clean, too artificial," says Eimhin McNamara, an animator at Cartoon Saloon - which produced Wolf-walkers. He also shares that "it helps you feel like you're actually in the head of the character." To help with the time-consuming process, they played with the frame rate - dropping it down from 24 fps (frames-per-second) to 12 fps, meaning they drew 12 images for every second. Still a very painstaking task if you ask me!

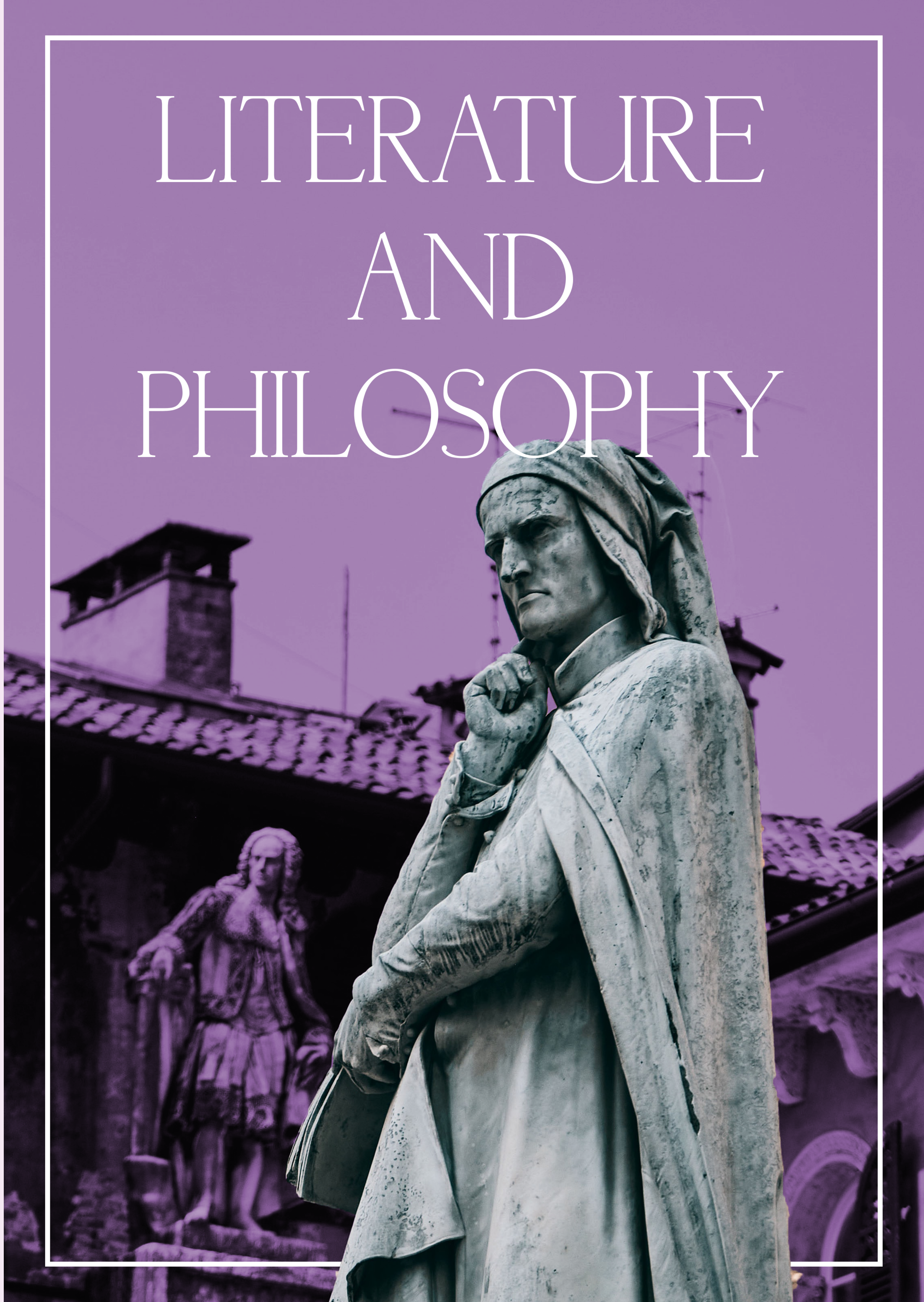
Talking more about the video games side of things, pseudo 3D games first came about in arcades in the mid-1970s. 'Interceptor', a combat flight simulator game by Taito, was supposedly the first arcade game to make use of 2.5D with its "eight-way directional movement, shooting down enemy planes, and depth perception by size-changing enemy planes." Moving further in time, we find more titles such as Donkey Kong Country, which features 3D characters and backgrounds, but with restricted 2D plane movement - ending up as a 2.5D game.



There is however, a key difference between 2.5D used in games and in films. In video games, this technique is used more often by playing with perspectives. An isometric view, such as the one depicted on the right, gives the illusion of 2D characters being shown as 3D. Nevertheless, like Donkey Kong Country, the side-scrolling 2D plane movement makes games like these in 2.5D perspective. The reason for using this technique was initially to do with limitations of the hardware systems - e.g. generating full 3D games required a lot more power than a NES could provide at the time. Nowadays, much like with films, it's a choice made for the realism and in-depth, immersive feel you get from a 3D-style.

In conclusion, 2.5D is a relatively generic term that describes a combination of animation styles used in making films and video games or an illusion of 2D and 3D. In fact, most of the entertainment that you usually consume is in fact in 2.5D - you just didn't know it! Keep an eye out for it next time you watch a Netflix movie thinking it's entirely either 2D or 3D, looks can be deceiving...

LITERATURE AND PHILOSOPHY



Can science tell us how we should live?

Audric Thakur

Part 1

'Science', as is framed in the question, is a slightly nebulous term, so I'm interpreting it to mean 'the use of an objective scientific method'[1]. Therefore, when I say, 'the science has told us not to eat McDonald's', what I mean is that 'multiple concordant studies in recent times have concluded that eating McDonald's out of moderation will result in long-term weight gain'[2].

While that makes sense, it does include a quick sleight of hand on my part which is worth exploring. In the second statement, I assumed that the implicit objective of the first was to maintain good health. The phrase 'Should I eat McDonald's?' by itself has no aim contained within it, so technically speaking, we can't answer.

While it holds colloquially, even if I had asked, 'Should I eat McDonald's to avoid gaining weight?', it's still difficult to reply in a yes or no fashion. A rigorous answer would take into account genetic factors; the other parts of your diet and lifestyle, specifically how much McDonald's; and so on to ultimately determine whether you would gain weight.

Therefore, every time we use the word 'should', it's necessary that we supply an objective – i.e., a larger aim within which the action I 'should' or 'shouldn't' take lies – and that we supply the necessary context that would apply to the question.

It's necessary to clarify that by saying '...how should we live?', I'm assuming that 'live' can be broken down into a series of more specific verbs. It's more productive to consider 'how should we live?' to mean how we should act in a finite and limited situation – and then assume that 'life' as a whole refers to a series of overlapping finite situations – rather than by trying to wrestle with fundamentally how we should live our entire lives start to finish in one go. We might end up there but it's not a useful place to start.

It's also important to clarify why I'm able to use the phrase 'the science' so easily, and why I intend to for the rest of the essay. As mentioned, I'm interpreting 'science' to be a method that ensures objectivity and lack of error. Therefore, 'science', in any capacity, can refer to both the existing research of our time – i.e., currently published work on the health effects of McDonald's – as well as the hypothetical research we haven't taken yet – i.e., the implications of McDonald's that we're yet to discover. After all, both of which still fall under the domain of science, given that the correct method is used. Therefore, it's possible to say a phrase like, 'although science can't tell us how we should act right now, if we had the right technology and method, then it could'. And given that 'science' in this case refers simply to a universal method – to quote Carl Sagan[7], "Science is more than a body of knowledge. It's a way of thinking, a way of sceptically interrogating the universe with a fine understanding of human fallibility," – we should be able to consider both existing science (research we have already taken) and potential science (research we either haven't done but could or research we will only have the means to do in the future) with equal merit.

Back to McDonald's.

Following our example, if I asked if eating a McDonald's takeaway (of specified caloric value and chemical makeup), providing the context of my diet and genetics and so on, would make me put on weight in the form of stored fat, it's reasonable to say that it's within the reach of science to answer that question. We either already have, or could obtain, the data to answer conclusively, yes or no. And we can reframe 'if I eat at McDonald's, I will gain weight' to fit into the context of the question by adding the modal verb 'should', as we discussed before. Therefore: 'If I want to maintain my weight, I should not eat at McDonald's'.

As demonstrated, 'science' has told me how I should act in that situation. As such, when I supply an objective and obtain enough data, then yes, science can tell me how to live.

Part 2

You might argue that such a statement can't be concluded from a single handpicked example. And you're right. I specifically chose an example which involved an easily quantifiable value – whether the numerical value of fat in my body would increase after regular McDonald's consumption, or whether my daily processes would be sufficient to burn those calories off after each meal. When the aims and objectives we have – to pass a test with a certain score, to achieve an income above a certain value – the metric for success is black and white. It's in these cases that science can tell us how we should live with total certainty.

This includes cases that aren't so obviously numerical. If my objective is for a ball in my hand to bounce, I can quantify this as whether the height of the ball post-bounce is of similar height to pre-bounce, at which point we can conclude that the ball 'bounced' (success). Any height lower than that, we can define as the ball having 'dropped' (failure). In this way, most isolated examples can be quantified, and by extension, used in a scientific context.

Let's try, 'What should my favourite colour be?'. You might instinctively consider this to be a subjective and personal question, but I would argue that it too can be quantified. Current research suggests that the human eye perceives around a million different colours (since we have 3 types of cone cells each perceiving 100 shades)[3]. Let's give each colour a score from -100 to +100 based on how strongly my brain reacts to seeing that colour. For this, we can use a machine that detects levels of electrical activity in the brain.

Since there are a finite number of colours, it's reasonable to assume that there will be a single colour, i.e. aquamarine, that reaches something like +99.82, beating out bright orange at +99.81, making it objectively my favourite colour. (There will be colours lower down with identical values, like a bell curve).

There are two ways to interpret this information. It's true that, with a microscopic enough scale and the supremely rigorous scientific method, every example can be made into a quantifiable measure of success or failure. With the example of colour, a common objection is that there are an infinite number of colours, and so there should be a smaller subset of infinite colours all scoring +100, at which point science can't give me a single highest scoring colour; the scientific method has fallen short. But this objection fails because the human eye can only perceive a finite number of shades. The fact that our experiences are limited by our body means that at some point, all data is discrete data with a high enough resolution.

But the second conclusion that I draw is more cautious. It reiterates that the assumption I took in the beginning that 'science' refers to both existing and hypothetical data is simply an assumption. While it's interesting to think that, at some level, science really could tell us how we should act in every isolated situation, the question asks 'Can science...?', not 'Will...?' or 'Could science...?'. As with every generation, our science of today is limited, so when we disregard the potential of the scientific method, there are cases where contemporary science just can't tell us how we should live. Moreover, just like how we used to believe that the universe was geocentric[4] or that cigarettes were beneficial to our health[5], there are cases today where we as a scientific community are wrong and don't know it yet. To me, this produces a more thought-provoking conclusion: for many of us, science does indeed inform the way we live, even when it is wrong.

Part 3

Things have been looking pretty favourable for science so far – it appears that in short, the answer to our question is simply 'yes; see terms and conditions'. And that's been true so far. And yet, hopefully, by the end of the essay, you'll begin to see the answer as a spectrum of many answers.
Firstly, some clarity is required on what an 'isolated example' really is. I've used the term now and then to break down how we should 'live' into how we should 'act in this situation'. As I said, it's easier to grapple with. Specifically, I use it to refer to all of the actions that are taken to achieve a single objective. When my single objective was to ascertain my favourite colour, I undertook a series of actions, all of which I consider to be part of one isolated example.

The difficulty comes when trying to ground these situations back into reality; going back from individual examples to 'living' as a whole. The difference is that when I have a singular objective in mind, there are a finite number of ways to achieve it, of which there is a single 'best' way to do so, i.e. most efficient and effective. I briefly mentioned an example of achieving above a certain score on a test. Either through existing scientific evidence[6] or by conducting new experiments and evaluating the data, we can determine the single most effective set of actions to take to achieve a particular score. Simultaneously, if I have responsibilities as a single parent to keep my child emotionally and physically healthy, given the nature of her genetics and lifestyle, and so on, there are a finite number of 'best' ways to achieve that, too.

But what happens when I try to achieve them at the same time? Within the space of a day, let alone a year or a lifetime, each of us has hundreds of major to minute priorities that overlap simply by coexisting in my life at the same time. And the more objectives I have, the more potential outcomes there are that day for how things eventually turn out. Since I can't complete every objective to the same degree, the varying degrees of moderation to which I can achieve each task gives rise to an infinite number of possible outcomes, of which some are better than others. However, just like how increasing the number of colours I could choose from means that there will eventually be multiple colours that are equally scoring, there will necessarily be multiple 'best case' outcomes that just happen to prioritise different objectives over others.

See, to say I should do something (to achieve an objective) implies that there is a single best way to go about it. If there are many equally viable actions I could take, there is no scientific imperative to choose one of them over the other, and as such, no single action I 'should' take. As such, the more equally viable actions that arise, the lesser the role of science is in determining which I should choose.

This is why I think that our answer should be a spectrum rather than a clear yes or no answer. The fewer objectives we have at once, the greater the likelihood that there is one single best compromise that most effectively utilises our resources in achieving as many objectives to the greatest degree. However, as we juggle more and more objectives, the likelihood that there's more than one 'optimum' course of action determined by a scientific method increases. Simply put, the more colours we have to choose from, the greater the chance that ten, twelve or a thousand of them will all score 99.99. And out of that group, there's just no scientific method to choose the best colour; the utility of science decreases as the number of objectives you have increases.

After all, to quote The Cincinnati Enquirer, 'The saying 'Life is just one damn thing after another,' is a gross understatement. The damn things overlap.'. Unfortunately, the more damn things you have, the less that science can help.

Isekai Light novels to literary theory?

Jim
'Rujing' Li

The isekai genre remains one of the most popular genres in web novels, light novels, manga and anime. As of August 2022, Anime Planet lists over 205 anime as having the isekai tag; over 1877 manga as having the isekai tag; and novel updates lists over 2100 light, web or published novels as having at least one of the following tags: transported to another world; transported into a game world; reincarnated in another world; and reincarnated in a game world. The origin of the isekai can be traced to the Japanese folk tale of Urashima Tarō, who saves a turtle and is brought into a magical undersea kingdom. After spending what he thinks is 4-5 days there, he returns to his home, only to find that 300 years have passed. That is what an isekai is: it is speculative fiction (fiction that incorporates elements not grounded in reality, recorded history or our universe) in which one or more protagonists are transported (usually through being summoned or dying and then reincarnating) into another world. From there anything can happen as the very nature of the setting allows the author to do whatever they want. The term isekai itself (meaning "other world" or "different world" in Japanese) was used sporadically until 2013, after which a substantial increase in searches can be observed (see fig 1).

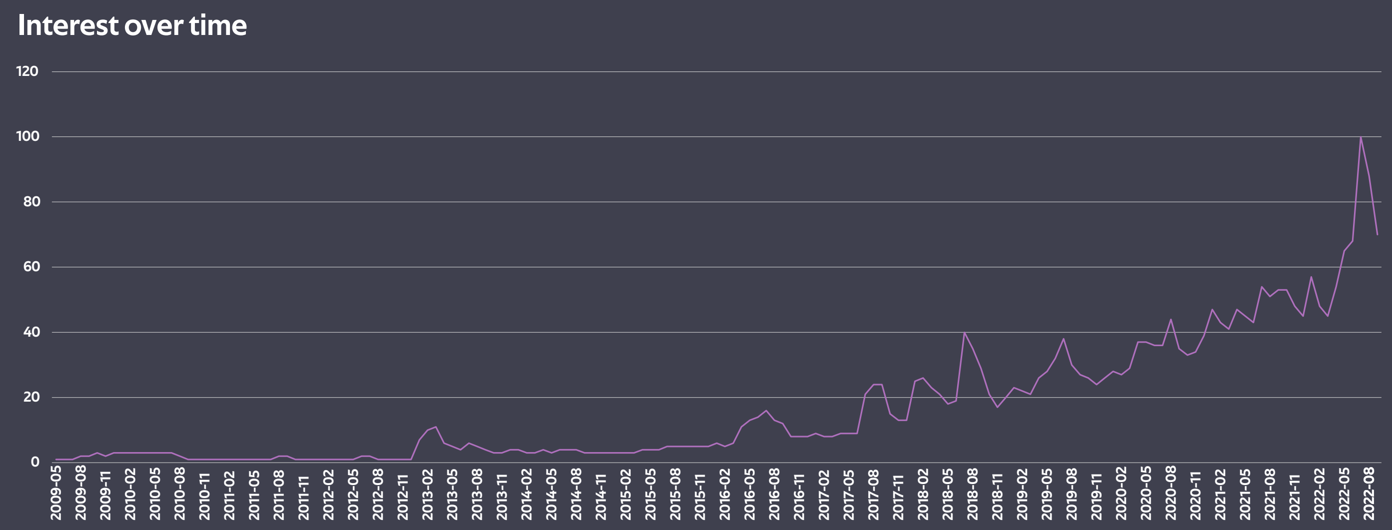


Fig 1. Google trends data showing frequency of the use of the term "isekai" in searches from 2004 with the data normalised to search frequency in July 2022. Frequency of use in July 2022 is arbitrarily set to 100.

With isekais having such a specific set of boxes they have to tick to be classed as an isekai, it is not surprising that Azuma Hiroki's grand nonnarrative can be applied and can also be commonly seen in all modern, mainstream isekai. Azuma's grand nonnarrative suggests that there is no grand narrative but instead all stories consist of tropes and clichés from a canonical pool of such recognisable to the reader. This collection of tropes and clichés is the so-called grand nonnarrative. It is blindingly evident how this applies to modern isekai. After all, they all feature one or more protagonists being transported one way or another to another world or setting; a lot of them receive new skills or "cheat abilities" (read legally distinct plot armour); and usually there is magic and monsters. More niche tropes include the "game system" (where the world the story takes place has a system similar to that of role-playing games that has a level up mechanic etc); the death-by-truck-leads-to-reincarnation, which has spawned the meme of "truck-kun"; and more fantastical elements like the adventurers' guild or the summoned protagonist(s) being chosen hero(es) and whatnot. Dr Paul Price, in his essay "A Survey of Story Elements of Isekai Manga" asserts that these stories "have no higher meaning other than they are created with the elements and characteristics that come from the grand nonnarrative". Whilst this can be seen to be true with a lot (and I mean the vast majority of isekais you read), it obviously does not apply to all isekai stories and there are some that subvert certain clichés and tropes to great effect but it then poses an interesting question: is it (the story) still part of the grand nonnarrative if it uses elements not part of the so-called canon? Or rather, if a subversion of an isekai trope (such as the protagonist being summoned to another world to be their saviour) is successful the first time (turns out the protagonist was mistakenly summoned alongside the saviour and is unwanted!) and then is copied ad nauseum by other prospective isekai stories, at what point does the subversion of the trope/cliché become one itself and become a part of the grand nonnarrative canon? Also, what even defines the canon of the grand nonnarrative? Is it merely just elements and characteristics of stories that are instantly recognisable to the reader? If so, isn't everyone's reading experience different?

With such confusing queries, it does seem that the isekai genre itself can offer a solution, or rather, the start to one. After all, isekai by the very nature of its being and creation is recognisable. Does your story involve being transported to another world? Isekai. Does your story not involve being transported to another world? Not an isekai. But is that really all that makes up an isekai? No matter how many tropes you tack on to your story, without the crucial "character is transported to another world", your story is not an isekai. So is that it? Is being transported to another world the only characteristic of an isekai story necessary for it to be called an isekai? Yes, it appears to be and that is why isekai is such a popular genre.

The thing with isekai is that the otherworld can be whatever the author wants it to be. As such, you can tailor any other genre to also be an isekai. Do you want some magic in your slice of life romantic comedy? Boom, make it an isekai. Do you want your coming-of-age story to happen in an interesting setting and not Earth? Boom, make it an isekai. Do you want to exposit all the cool and exciting lore of your magical world but can't think of a way to do it subtly and/or naturally? Boom, make it an isekai so your clueless, out-of-this-world (literally) protagonist acts as the perfect catalyst for any exposition dumps you want to do to slow down the plot! With so much variety and with how saturated the isekai genre is nowadays, there seems to be an isekai for every niche. Want to read about how a guy got reincarnated as a sentient sword and is wielded by a cute cat girl? There's an isekai of that (Tensei Shitara Ken Deshita). Want to read about a guy got reincarnated as a vending machine and now travels with a cute girl? There's an isekai of that (Jidouhanbaiki ni Umarekawatta Ore wa Meikyuu ni Samayou). Want to read about how a guy got reincarnated as a slime, makes friends with an ancient, powerful dragon and leads a village of goblins to creating a global superpower? There's an isekai of that (Tensei Shitara Slime Datta Ken). Naturally, having so many isekai stories out there means that a whole lot of them are garbage and very few are good (and tasteful) pieces of literature.

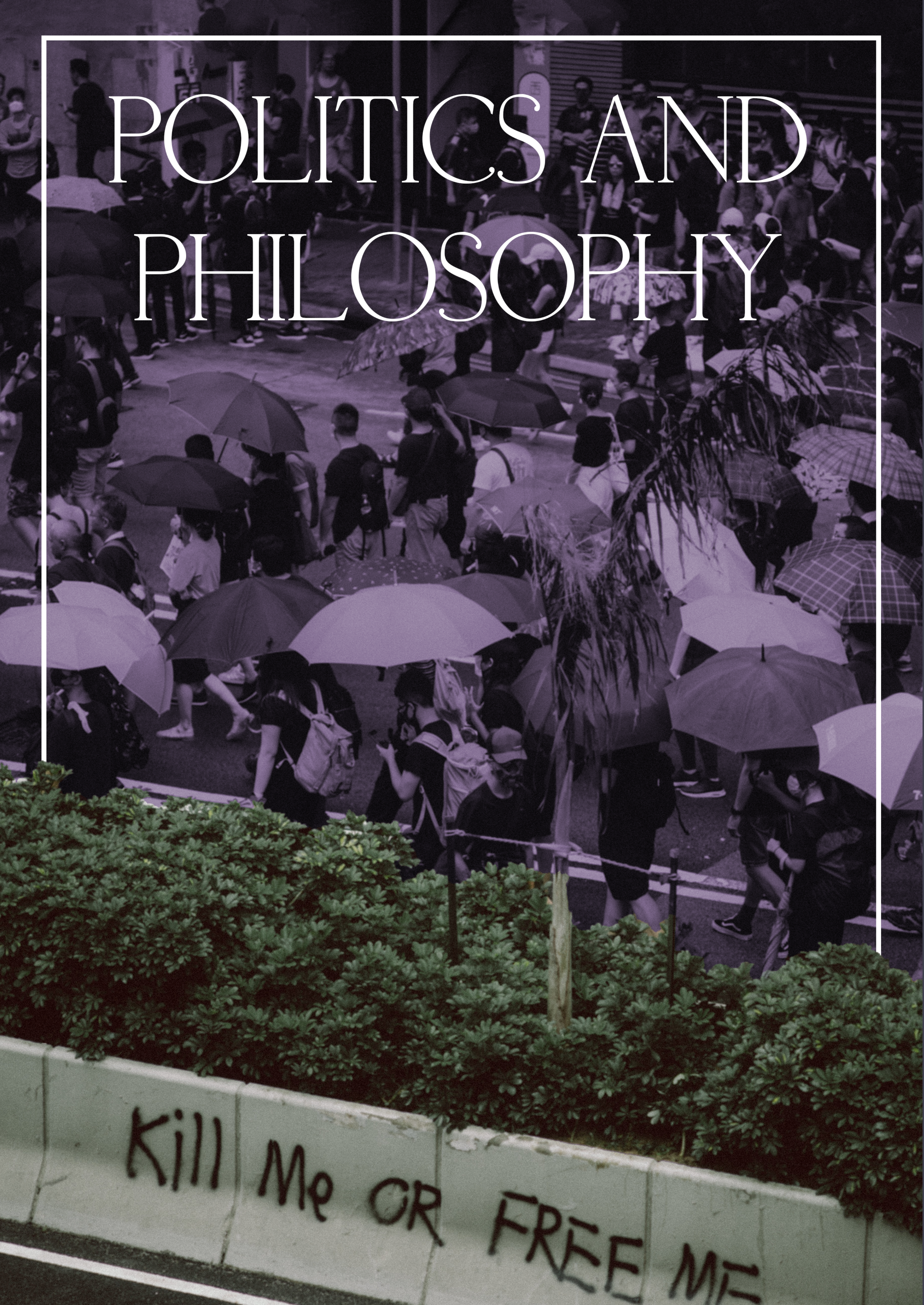
There are certain memes surrounding the isekai genre in popular culture. The aforementioned meme of "truck-kun" is perhaps the most prominent, caused by the trope of having your isekai protagonist reincarnate as a result of being run over by a truck - either because the driver was drunk or asleep at the wheel and unwittingly ran the unlucky (or lucky?) protagonist over or because our intrepid, heroic protagonist pushed some baby/child/teenage girl/ old lady out of the way of a truck whose driver was drunk or asleep at the wheel (wow, misandry much?). Another meme is the absurdly long title, prominent in mostly Japanese isekais (though not limited to just isekais, it seems to be a phenomenon affecting all genres), and acts more like a blurb than the title of a book. The most damning is perhaps the meme that all isekais are self-insert power fantasies of the author, characterised by an overpowered (male) protagonist with as much personality as soggy cardboard who inadvertently gains a harem of beautiful women in another world. This brings up a problem with the isekai genre in the modern age, an age where anyone with access to the internet realise their wildest fantasies onto the digital page and subject such horrors to any hapless netizen. The problem with isekai is also isekai's greatest appeal and has been established previously in this article. It is the problem of the author being able to write whatever they want with a character that is innately somewhat relatable (after all, the character that is transported is our only link to the world we are familiar with so we feel a kinship with the character as the only remotely relatable thing in this new world). This leads to the trash of the isekai genre: the drivel perpetuated by incels, designed only to act as a way to hide their insecurities behind the generic face of the embodiment of their power fantasies. Let this not, however, distract from the truly wonderful stories that the isekai genre has to offer, who shine even brighter when compared to the relative trash littered across the internet.

Similar to the grand nonnarrative proposed by Azuma, narrative consumption, as proposed by Ôtsuka Eiji, describes the act of reading smaller narratives not only for direct personal enjoyment but also as a way to access a larger, grand narrative. This grand narrative is the framework for smaller stories and is only accessed by the consumption of smaller narratives, hence narrative consumption. As I perceive it, the truly immersive, well-written isekais are ones that offer smaller narratives as a way to access the larger plot whilst also keeping to (or cleverly subverting) the characteristics and tropes of the grand nonnarrative. The blend of both allows the author to fully explore the world they have created without having all the exposition and lore dumped onto the main protagonist which slows down the plot. Having multiple narratives and viewpoints resolves that problem and furthermore, it develops the side characters alongside the main character as you get their thoughts and opinions on the various events of the story - it can also be used to progress the plot from another point of view or an interesting angle. That being said, isekais are written to revolve around the transported character and as such, they must be the focus of attention in most of the story otherwise the story may as well be a fantasy story without the added element of the transported character. The use of multiple narrative perspectives just allows deeper development of the world and the characters that inhabit the world rather than the reader only getting the narrative perspective of the main character, which limits the amount and also what sort of exposition or lore they can be exposed to. There are several great examples of this and each of them is a powerhouse in the isekai genre however, my favourite example is "So What If I'm a Spider, So What" ("Kumo Desu Ga, Nani Ka?").

[MINOR SPOILERS AHEAD] This story subverts a lot of the common tropes of the isekai genre: the protagonist is killed in a mysterious explosion alongside her classmates and is reincarnated as a spider monster without any cheat abilities in a perilous situation (the bottom of a dangerous labyrinth, which actually adds stakes and tension to the story), thus subverting the tropes of being reincarnated as a human or humanoid and receiving some "cheat ability" to boost them ahead of their peers. Through various sidestories you are drip fed the events of the outside world through the perspective of her other classmates (also reincarnated but more in line with what is expected of an isekai) which succeeds in developing the plot and world of the story as a major motivation for one of the characters is to reunite the class which involves finding our protagonist who's stuck in the most dangerous place on that world. Through the side characters the author, Okina Baba, also subverts the hero trope of isekais (where a transported character is the chosen hero who will save the world) by having the hero be an existence that is actively harmful to the survival of the world and ALSO offers an interesting spin on the perception of time in the story. By having the side character chapters appear so early in the novel, you, the reader, are let to believe that the events of both the side character and protagonist narratives are happening in tandem; however, it becomes subtly evident that they are not - the main events of the story (from the protagonist's perspective) are happening a full decade before the events of the side character stories and acts as a satisfying payoff to the small details strewn about the earlier chapters that lead up to this revelation. It successfully blends both the grand narrative and the grand nonnarrative to achieve a story that clearly revolves around the core cast of transported characters whilst using the smaller narratives to fully establish all the key character motivations, backgrounds and lore of the world as well as drawing from the pool of isekai tropes to deliver a fresh take on the isekai genre and a unique story.

The isekai genre is expansive and limitless in the stories it can tell. The worlds that are introduced can create such a feeling of wonder and desire to know more that it can suck you in, much like the protagonists of the stories themselves. It's easy to get lost in the sea of isekais, in the characters and in the plot - something that the authors themselves seem to be victim to. After all, how many isekais have completed stories? It doesn't seem like a lot of them do: 383 of the over 2100 web, light and published novels on novel updates are listed as completed and that's just the translations of the original texts. Whilst isekai is mostly lighthearted genre consumed for mere pleasure, a bit of light reading to do when you're bored, it is also a lesson in the dangers of burn out and capitalism. The isekai series "Overlord" ("Ōbārōdo") has been discontinued due to the author, Kugane Maruyama, feeling burnt out after having written the series since 2012: a series that has seen 16 light novel volumes, 17 manga volumes and 4 anime seasons. Many a web novel isekai once given a manga series or anime or official publication as a light novel seems to be less interested in the overall story, the grand narrative, and more on making money. My beloved "Kumo Desu Ga, Nani Ka?" is a victim of this - the ending of the web novel was suspended whilst the author wrote the light novel (the official publication of the story) and whilst the anime was in production. When the ending was finally delivered, it was lacklustre to say the least. So whilst Dr Paul Price's comment that these stories "have no higher meaning other than they are created with the elements and characteristics that come from the grand nonnarrative" can be held as true for the plot, perhaps they hold meaning as cautionary tales against burning out doing something you find passionate or forgoing good storytelling to make more money.

POLITICS AND PHILOSOPHY



Morality of acts involving only consenting adults

EG
Elijah Gibbons

I – Authentic consent and the problem of consent to harm

Before considering how consent functions in moral cases, it must first be articulated what consent actually is. Basically conceived, consent is “an act of permitting something to be done or of recognizing some authority”[1]. In limiting our question to concern of consenting acts involving adults though, we tacitly recognise some limits on consent differentiating it from mere assent. We rightly restrict children to being incapable of consent; in On Liberty, Mill writes that personal sovereignty (and therefore consent) is the preserve only of “human beings in maturity of their faculties”. By the same principle, we might similarly deem those with severe mental health conditions to be incapable of consent[2], or those whose faculties are only temporarily impaired, for example, if they are drunk. Clearly, there is some threshold of freedom and rationality before which an agent’s consent is inauthentic.

Mill justifies that it is right to restrict someone’s capacity of consent in cases where they “must be protected against their own actions”, implying that, upon the recognition that consent is at least intended as a moral sanction for action, there is a need to avoid sanctioning damage resulting from some irrational agent to their own harm.

Such an analysis immediately brings into question though whether an adult – specifically one who we would normally consider to be free and rational – can be deprived of consent on the same basis as children though, given that generally rational adults often do consent to their own harm. This problem will form the basis of the objections discussed in this essay whereby the view that consent is capable of absolving wrongdoing may be rejected - the problem that substantially rational adults are demonstrably capable of consenting harms (or at least to minor harms) to themselves.

There are certain cases, such as rape, in which the harm involved is itself dependent on consent, and which consent would substantially transform or annul. However, there is a majority of harms a person might consent to (for example, consenting to being hit) in which consent does nothing to alleviate the harm experienced in the act. At least according to a pointedly consequentialist view of ethics then, there is a justified default position here that consent does necessarily hold any moral significance. Acts involving only consenting adults may be morally wrong under the same conditions as any other act – the condition that they incur more moral harm than good.

To argue the contrary – that consent can annul moral harm – we require some more general principle of liberty, which defends a fundamental right for someone to act upon her own consent, and allow others to do the same to her also upon her consent; where she can decide which ethical duties apply to her.

II – The harm principle

Mill presents this kind of right in his ‘Harm Principle’, which states that “the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant.”. This principle, at first, seems to arise in respect of a fundamental concept of personal liberty, about which Mill writes “In the part which merely concerns himself, his independence is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign”.

As personal liberty entails that our sole moral limit is in acting upon other people, consent is defended as absolute, and annulling culpability for moral harm, both in the case of harming one's self, and in the case where one devolves their allowance to harm themselves to another in accordance with their "free, voluntary, and undeceived consent and participation".

There is a problematic sense though, in which the Harm Principle is both a fundamental expression of liberty, and simultaneously not about liberty at all. Mill's real justification for the absolute moral legitimacy of consent is not on the basis any "abstract" libertarian right, but rather given a utilitarian slant which makes liberty merely instrumental in achieving wider utilitarian ends (principally avoiding governmental tyranny). Regardless of whether the Harm Principle is justified as a political right – where politics may be taken as ethics in the context of society as a whole - there is a clear disconnect between such ideas and the purely ethical realm. From Mill's own consequentialist perspective, it is still the case that action between consenting agents may be wrong, provided that it is sufficiently isolated from wider society, and therefore Mill does not actually defend the moral impunity involved in acts of consent on an absolute basis.

III – Kantian liberty: the sovereignty principle

At the core of this tension whereby the Harm Principle functions best for political but not ethical legislation is also Mill's "negative" concept of liberty – a concept of liberty as deriving from the absence of external limits on a person's individual powers. While there is a strong case that freedom (and the value of consent) must be defended from external limits in any concept of liberty, the concept fails unless this is done on the basis of a positive libertarian axiom, not just under the threat of the tyranny brought about by abandoning liberty, as Mill expresses. This negative basis ultimately dissolves into a fundamentally anti-libertarian, consequentialist ethical system. To form a coherent basis for universally justifying consent, liberty as such must manifest as both a fundamental political and an ethical concept.

It is also through this inconsistent relationship with liberty that the Harm Principle, intending to justify the "wrongless harms" involved when someone consents to their harm, becomes vulnerable the problem of "harmless wrongs." In "Beyond the Harm Principle", Arthur Ripstein's argument against Mill is structured around the example of a person who invades someone's home and takes a nap there, without damaging the home or tangibly setting back the owner's interests in anyway . In this case, there clearly is some principled moral wrong involved, associated with a violation of the homeowner's liberty without consent. However, the Harm Principle fails to legislate against this on account of its utilitarian basis: it must hold that there is no harm to the homeowner, so there is no wrong. Generally, the Harm Principle suffers on account of being incapable of prohibiting any violation, so long as it does not actually manifest in harm.

In recognition that avoidance of harm cannot be an exclusive basis for wrongdoing, Ripstein exchanges a utilitarian justification for liberty with a Kantian one, in which freedom is grounded in the inherent dignity of the human as an end-in-itself.From this he derives the "Sovereignty Principle", which states that "The only legitimate restrictions on conduct are those that secure mutual independence of free persons from each other". Through this principle, harmless wrongs such as the home-invasion are successfully prohibited on the basis that the wrongdoer has placed the victim (or, in this case, the victim's property, over which they are also sovereign) as a mere means, violating the independence and sovereignty of the victim.

This "Sovereignty Principle" provides a much stronger basis for moral impunity in the case of consenting harm. An individual's sovereignty entails that they can choose, via consent, which ethical duties apply to themselves. Therefore, whenever an agent consents to something done to them, what is done to them is in accordance with their choice and their interests, and therefore their dignity as ends-in-themselves, and their independence is never violated, regardless of the context of the specific harm which troubles Mill's principle.

When they consent to their own harm, "any harm that ensues is not an interference with sovereignty. Other harms do interfere with sovereignty, but it is that interference, not their harmfulness, that merits prohibition".

IV – Consent in violation of human dignity

While this concept that one's dignity can never be infringed upon according to consent seems intuitively reasonable, there is an important objection in the case of particularly grave harms. This is not merely on the consequentialist basis articulated in the introduction to the essay, although it could be argued that the extent of the harms effected by certain consensual acts might be so great as to exceed the value of maintaining any principle of liberty. Instead, there may be cases in which some harm is so extreme as to infringe upon a consenting subject's liberty, and further than this, their human dignity itself, purely by virtue of the gravity of the harm.

Ripstein only entertains that freedom can be wrongly restricted, and human dignity degraded as the result of one actor enforcing their will upon another and the things over which they are sovereign. However, in "The Moral Limits of Consent as a Defense in the Criminal Law" Dennis Baker argues a harm like death can clearly destroy our human dignity in of itself to a far greater extent than anyone else's decision to use us for their own ends, and it is because of this that consent is an unreasonable justification for murder.

In his articulation of the humanity principle, Kant himself explains that we are "not at liberty to dispose of that humanity which constitutes my person, either by killing, maiming, or mutilating it," on the basis that even the subject of a harm is capable, in consenting to the harm, of treating herself as a mere means and thus, infringing upon their own sovereignty.

In response to this, Ripstein simply recognises that the Sovereignty Principle has "no resource to prohibit suicide, as it does not involve domination", providing only an argument against consent to murder by an external agent. However, both the problem of suicide and the problem of consent to one's own murder are of equal significance from this Kantian perspective as, in moving from suicide to consenting murder, while the structure of culpability might be altered, it is still the same ultimate ethical rights and consequences which are at stake.

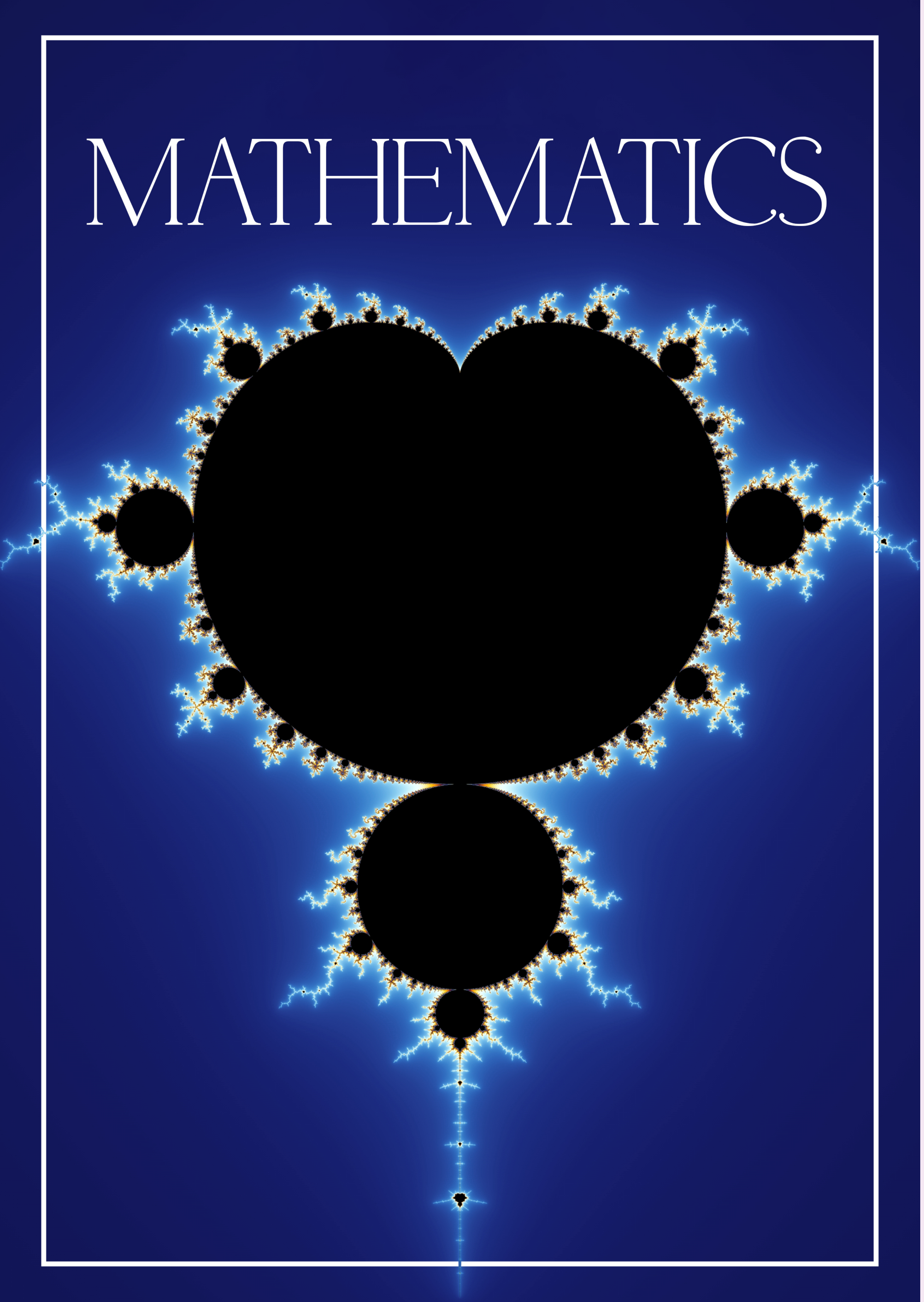
It is my view that Ripstein misses a significant objection to this though, which is that Kant and Baker, in arguing that an agent is capable of consenting to the degradation of their own dignity, misrepresent what constitutes authentic consent. Particularly, they do not rightly consider whether such agents satisfy the condition of rationality as established at the beginning of this essay. For example, Baker recognises that we have an inalienable right to life and yet he conceives of a rational agent could deny this right while maintaining the rationality needed to be capable of consent. I believe it also more widely evident that, in any of the cases Kant presents – the desire to kill, maim, or mutilate one's self – such desires are always the consequence of some mental disorder.

Obviously an agent's requirement for rationality for consent is not absolute; there is irrationality which may be permissibly involved in any exercise of personal autonomy. But there is a central and dangerous irrationality necessary for them to degrade their own dignity involved when one treats one's self as a mere means, forgoing any ethical duty to their own life, and general humanity, which certainly precludes them from consent.

Ultimately, if the purpose of consent is to represent one's fundamental interests, and an agent's interests fundamental are not being represented in the act they consent to (i.e. in one in which they treat themselves as a mere means), it is clear consent has been misattributed.

V – Conclusion

In general then, while it has been shown there is a valid defence of consent as providing moral immunity via the positive, libertarian basis of the Sovereignty Principle, we must also recognise there is an extent to which this libertarian ideal restricts consent; it introduces a rational threshold through which we must pass to gain it. On the basis of the Sovereignty Principle then, we can conclude that the acts of consenting adults are never morally wrong, so long as we are aware that consent is held only on the condition that we do not abuse it in degrading our own dignity.



Karthik Donthula

Mathematical chaos and chaos theory

Chaos is when the present determines the future, but the approximate present does not approximately determine the future. Chaos theory is the study of apparently random or unpredictable behaviour in systems governed by deterministic laws.

What is chaos?

If I drop a ball on the ground, I can give a reasonably accurate estimate of where the ball will land. My estimate however will not be exact to the nearest micrometre, and the main reason for that is due to not knowing the exact starting point of the ball when I dropped it. There is some level of uncertainty in my estimate, but it is of an acceptable amount. Most importantly, the level of uncertainty in my estimate equals the level of uncertainty in the starting conditions.

If I drop a ball on the edge of a knife however, and I want to know which side of the knife the ball lands on, the exact position the ball is released from is significantly more important. A small variation in starting position can change the outcome entirely. If I can't tell which side of the knife the ball is released from, I can estimate the probability of each outcome and they're 50% each. There has been an increase in the uncertainty of the outcome.

Finally, I'll drop a ball onto a peg-board. At every row, the ball may fall on either side of that peg. The route that the ball takes and its final destination depend precisely on the starting conditions now. Mathematical chaos resembles an infinitely long peg-board. If there is any uncertainty at all about the starting position of the ball (and in any real world scenario there will definitely be at least some uncertainty), then the route of the ball remains unknown to us. Even the tiniest of changes will fundamentally change the course of the ball. This means that with the slightest of uncertainty, we will be totally oblivious of the outcome. This is what mathematicians call chaos.

The simplest chaotic system

Now we know what chaos is, we can start exploring some chaotic systems. Chaos can originate from the simplest things, and we will now be looking at the simplest chaotic system: the logistic map.

Let's take a fixed variable r, a variable x (that is between 0 and 1) and another variable (1-x). Now, we can form a simple sequence with $x_{n+1} = x_n * (1-x_n) * r$. Let's first take r = 2 and $x_0 = 0.25$, meaning $1-x_0 = 0.75$.

$x_1 = (2)(0.1)(0.9) = 0.18$

$x_2 = (2)(0.18)(1-0.18) = 0.2952$

Next, you'll get values of $x_3 = 0.41611392$, $x_4 = 0.4859262512$ (10dp), $x_5 = 0.4996038592$ (10dp), $x_6 = 0.4999996861$ (10dp). As you can see, the sequence is getting ever closer to 0.5, which is called the attracting point of the system. The starting value of 0.1 has no relation to the attracting point, as with any starting number between 0 and 1 will yield the same attracting point.

What's interesting however, is what happens when we change the value of r. If we change r to 3.1, keep the starting value as 0.1, and round of our results at 2dp, the sequence will go:

0.1, 0.28, 0.62, 0.73, 0.61, 0.73, 0.60, 0.74, 0.59, 0.75, 0.59, 0.75, 0.58, 0.76, 0.57, 0.76, 0.57, 0.76, 0.56, 0.76, 0.56, 0.76, 0.56, 0.76, 0.56.

This time, the sequence doesn't tend to 1 value but rather fluctuates between 2, namely 0.76 and 0.56.

This is called an attracting 2-cycle.

When r = 3.5, we get an attracting 4-cycle between 0.50, 0.87, 0.38 and 0.83.

We can figure out new values of r where we find attracting 8-cycles or 16-cycles, etc. This is called a sequence of bifurcations. Once r exceeds approximately 3.56995, something new happens. Taking r = 3.57 will give us the sequence 0.1, 0.32, 0.78, 0.62, 0.84, 0.47, 0.89, 0.35, 0.81, 0.54, 0.89, 0.36, etc.

It's easy to think that if you let the sequence run on for long enough, eventually, a pattern will emerge, but it will in fact not. This is the core of mathematical chaos. No matter how long the program runs for, the sequence just jumps around seemingly randomly. Furthermore, even if we change the starting value by microscopic amounts, the new sequence will look completely different to the original.

Chaos in real life

The (arguable) founder of chaos theory is Edward Norton Lorenz, a mathematician and meteorologist, who, in 1961, used a simple digital computer to simulate weather patterns. This simulation was modelled by 12 variables, such as temperature and wind speed. He wanted to see a certain sequence of data again so, to save time, he started the simulation in the middle of its course by entering a printout of data that corresponded to conditions in the middle of the original simulation. What was surprising was that the predictions made by the machine were completely different from the previous calculation. The reason for this was due to the printout containing rounded values. The computer itself worked with a 6-digit precision (0.991231), while the printed values contained 3-digit values (0.991). The difference is minuscule and the consensus at the time was that there should be no difference practically. Lorenz has just discovered that small changes in initial conditions produce large changes in long-term outcome.

This means that as the resolution of weather readings increases, the predictions of future weather will change drastically. In 1969, Lorenz described the butterfly effect which entails that even the tiniest factor has a significant impact on an outcome, such as a butterfly flapping its wings in Brazil being the cause of a tornado a few weeks later in Texas.

Let's take another field now, such as astronomy. In space, bodies, such as planets and stars, orbit around each other under the mutual attraction of gravity.

When there are just 2 bodies, their paths of orbit are extremely simple to calculate. Planets orbit stars in ellipses. Stars can be locked in orbit around each other. When two bodies move towards one another, if they travel slowly they will spiral inwards slowly until they collide or orbit each other indefinitely, and if they travel quickly they will fly along predictable parabolas or hyperbolas away from each other.

When there are 3 bodies interacting with each other however, the situation becomes very chaotic. The paths that the 3 bodies take are highly complex and are unique, so will never repeat unlike the clear ellipses or parabolas when 2 bodies pass by each other. The slightest change (e.g. one of the bodies arrives one millisecond later) will drastically change the outcome. For example, rather than performing a series of rotations around each other before flying off, they could now all collide together.

This is called the 3-body problem, and was the first discovered example of chaos in physics.

What is chaos theory?

Chaos theory is the branch of mathematics which focuses on finding underlying patterns and laws of dynamical systems that are highly sensitive to their starting conditions, that had previously been thought of as completely random, in states of disorder and with irregularities.

Chaos theory states that within the apparent randomness of chaotic complex systems, there are underlying patterns, interconnections, constant feedback loops, repetition, self-similarity, fractals, and self-organisation. These are the key principles of chaos theory.

The chaotic systems, like the logistic map which we discussed earlier, are predictable for a while and then 'appear' to be random. The amount of time for which a chaotic system can be predicted depends on 3 factors: how much uncertainty can be tolerated; how accurately its current state can be measured; and a time scale which depends on the dynamics of the system (called the Lyapunov time).

In chaotic systems, the prediction inaccuracy increases exponentially with time passed - for example, doubling the length of time passed in a forecast will result in an uncertainty which is more than squared. This means that an accurate prediction cannot be made for events too far in the future (after an interval of more than two or three times the Lyapunov time), which is when the system appears to become random.

A commonly used definition, made by Robert L. Devaney, says that to be classified as chaotic, a dynamic system must have three properties:

- It must be sensitive to initial conditions, which again relates back to the butterfly effect (a small change in one state of a system can result in large differences in a later state).
- It must be topologically transitive, which roughly means that neighbourhoods of points eventually get flung out to "big" sets so that they don't necessarily stick together in one localised clump.
- It must have dense periodic orbits, which means that there must be attracting points in the system, which we have briefly talked about earlier.

For any interested, I highly recommend you go fuel your interest further online as I have only briefly covered the main aspects of chaos theory in this article. Thank you all for reading, I hope you enjoyed this article and can now tell people the difference between normal chaos and mathematical chaos.

Lily
Hillman

Applications of advanced induction methods

Induction is a common proof method in A-level further maths. I aim to explain some important advanced methods not covered in the curriculum and their applications in proving significant mathematical theorems, including forward-backward induction, two-dimensional-dimensional induction, and more.

Introduction

Mathematical induction is a proof technique based upon showing that if a statement holds in one instance, then it must also hold for the next one. It can be likened to setting up dominoes so that if one falls over, then all of the others follow. Typically, this is used to prove statements of the form P(n) based upon a natural number n.

Some of the most natural applications of induction arise from spotting patterns in arithmetic. For example, one may consider the sums

$$\begin{aligned} 1 &= 1 \\ 1+2 &= 3 \\ 1+2+3 &= 6 \\ 1+2+3+4 &= 10 \\ &\dots \end{aligned}$$

and guess the pattern 1 + 2 + ... + n = ½n(n + 1). However, it may be more tricky to rigorously prove this result. This is where induction comes in.
Let P(n) be the hypothesis that 1 + 2 + ... + n = ½n(n + 1).
First, we establish a simple base case, such as n = 1: ½(1)(1 + 1) = 1.
Then, we show that if P(n) is true for some number n, then P(n+1) must also be true.
We do this by applying the result of P(n) to the sum 1 + 2 + ... + n + (n + 1):

$$\begin{aligned} 1+2+\dots+n+(n+1) &= \frac{1}{2}n(n+1)+(n+1) \\ &= \frac{1}{2}(n(n+1)+2(n+1)) \\ &= \frac{1}{2}(n+1)(n+2). \end{aligned}$$

This is just the statement of P(n + 1). In other words, if this works for one number, then it must work for the next. But we have already established P(1), which implies P(2), which implies P(3), and so on for every natural number.

We have now completed a proof by induction. As a note, the base case does not have to be 1: you may choose P(b) as your base case for some b, and then prove P(n) for all n ≥ b. If you wish, try proving the following theorem by induction as an exercise:

For all n ≥ 3, the sum of the angles of a polygon with n vertices is 180(n - 2).

Strong induction – the Fibonacci sequence

In the inductive step so far we have taken P(n) as given and then proved P(n+1). In fact, with base case P(b), we may assume all P(k) for b ≤ k ≤ n. For example, if P(3) is our base case, then P(3) implies P(4), P(3) and P(4) together imply P(5), and so on. This is known as strong induction, since it uses a stronger prerequisite assumption than standard ("weak") induction. One useful application of this technique comes from considering the Fibonacci sequence:

$$F_0 = 0, F_1 = 1, F_n = F_{n-1} + F_{n-2}$$

The first ten terms are:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \dots$$

What we seek is a formula for the nth Fibonacci number. Looking at the sequence, it is reasonable to guess that this formula is exponential in nature. Therefore, we substitute $F_n = x^n$ into the recursive formula to get:

$$\begin{aligned} x^n &= x^{n-1} + x^{n-2} \\ x^n - x^{n-1} - x^{n-2} &= 0 \\ x^{n-2}(x^2 - x - 1) &= 0 \end{aligned}$$

$$x = 0, x = \frac{1+\sqrt{5}}{2} = 1.618\dots, x = \frac{1-\sqrt{5}}{2} = -0.618\dots$$

However, x = 0 is clearly invalid for n = 2. We name the remaining two roots ϕ and ψ respectively, and guess that $F_n = a\phi^n + b\psi^n$ for some unknown a and b – in other words, a linear combination of powers of ϕ and ψ. Trying the first few values, we can update our hypothesis P(n) to:

$$F_n = \frac{\phi^n - \psi^n}{\phi - \psi}$$

We start with two base cases, P(0) and P(1) (i.e. n = 0 and n = 1):

$$\frac{\phi^0 - \psi^0}{\phi - \psi} = 0 = F_0$$

Now in the strong inductive step, we show that for any n, P(n−1) and P(n) imply P(n + 1):

$$F_{n+1} = Fn + F_{n-1} = \frac{\phi^n - \psi^n}{\phi - \psi} + \frac{\phi^{n-1} - \psi^{n-1}}{\phi - \psi} = \frac{(\phi^n + \phi^{n-1}) - (\psi^n + \psi^{n-1})}{\phi - \psi} = \frac{\phi^{n+1} - \psi^{n+1}}{\phi - \psi}$$

In the last step, we used the earlier formula $x^n = x^{n-1} + x^{n-2}$ for both ϕ and ψ. Now P(0) and P(1) imply P(2), P(1) and P(2) imply P(3), and so on, and the proof is complete. As another optional exercise, try to prove the following statement by strong induction:

Any natural number can be written as a sum of Fibonacci numbers.

Forward-backward induction

This is where our methods begin to become more esoteric and specialised. However, they are still useful tools for those who want to study maths at a higher level. In order to explain forward-backward induction, I will first present an example and then discuss the structure of the proof.

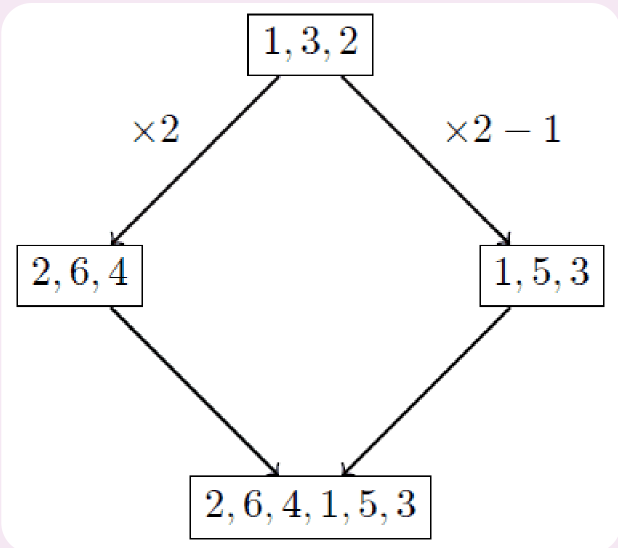
The problem is as follows:

Show that for any natural number n the sequence 1, 2, 3, … n can be rearranged such that the mean of any two numbers is not between those two numbers in the sequence.

For example, if n=4, the sequence 1, 3, 2, 4 would be valid, but the sequence 4,3,1,2 would not, as the mean of 4 and 2 (i.e. 3) is between them in the sequence. This problem would not be out of place in a maths olympiad. However, it may seem difficult to approach at first glance. One good strategy is to try small values of n and try to seek a pattern. For example, taking n = 6, we can construct a few sequences such as:

$$2, 6, 4, 1, 5, 3$$
$$3, 1, 5, 6, 2, 4$$
$$4, 2, 6, 3, 5, 1$$

One thing may stand out to you: the even numbers and the odd numbers seem to be separated. Why is this? Well, the mean of an odd number and an even number is not an integer, so it does not need to be considered. This neatly splits the sequence into two parts. In fact, given a sequence, we can multiply all the terms by 2 to get a sequence of even numbers, and take away 1 from each term in that sequence to get a sequence of odd numbers which both maintain our desired property. Finally, we can simply connect those two sequences to get a valid sequence with twice the length of the original.



So, if we take the trivial sequence for n = 1 (which is just the number 1), we can use this process to construct a sequence with length 2, then a sequence with length 4, length 8, etc. But how do we fill in the gaps between these powers of 2? Well, removing a number from a sequence is certainly not going to render it invalid. So if we want a valid sequence of length 13, for example, we can just construct a sequence of length 16 and remove the numbers 14, 15, and 16. Therefore, we have shown that a valid sequence can be constructed for any natural number n.

In general, a proof by forward-backward induction has three steps: establish a base case P(b). Then, show that for any n, P(n) implies some P(k) where k>n. Finally, show that when b<n, P(n) implies P(n−1). In effect, this is like skipping forward in large jumps, and then going backwards in small steps.

To end this section, I will present Cauchy's well-known proof of the arithmetic mean/geometric mean (AM-GM) inequality, which states that for non-negative real numbers a_1, a_2, … a_n, then:
Then:

$$\frac{a_1+a_2+\ldots+a_n}{n} \geq \sqrt[n]{a_1a_2\ldots a_n}$$

$$\left(\sqrt{a_1}-\sqrt{a_2}\right)^2 \geq 0$$
$$a_1+a_2-2\sqrt{a_1a_2} \geq 0$$
$$a_1+a_2 \geq 2\sqrt{a_1a_2}$$
$$\frac{a_1+a_2}{2} \geq \sqrt{a_1a_2}$$

The first inequality holds because a real number squared is always non-negative. Then, for the forward induction step, we show that for any n, P(n) impliesP(2n). (This is not the only possible way to "jump" forwards, but it is the most common.)

$$\begin{aligned}\frac{a_1+a_2+\ldots+a_{2n}}{2n} &= \frac{\frac{a_1+a_2+\ldots+a_n}{n}+\frac{a_{n+1}+a_{n+2}+\ldots+a_{2n}}{n}}{2}\\ &\geq \frac{\sqrt[n]{a_1a_2\ldots a_n}+\sqrt[n]{a_{n+1}a_{n+2}\ldots a_{2n}}}{2}\\ &\geq \sqrt{\sqrt[n]{a_1a_2\ldots a_n}\sqrt[n]{a_{n+1}a_{n+2}\ldots a_{2n}}}\\ &= \sqrt[2n]{a_1a_2\ldots a_{2n}}\end{aligned}$$

The first inequality follows from applying of AM-GM with n variables on $\frac{a_1+a_2+\ldots+a_n}{n}$ and $\frac{a_{n+1}+a_{n+2}+\ldots+a_{2n}}{n}$.

The second follows from applying AM-GM on the 2 terms in the numerator. For the backward induction step, substitute $a_n = \frac{a_1+a_2+\ldots+a_{n-1}}{n-1}$. Then:

$$\frac{a_1+a_2+\ldots+a_{n-1}+\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}}{n} \geq \sqrt[n]{a_1a_2\ldots a_{n-1}\left(\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}\right)}$$
$$\frac{a_1+a_2+\ldots+a_{n-1}}{n-1} \geq \sqrt[n]{a_1a_2\ldots a_{n-1}\left(\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}\right)}$$
$$\left(\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}\right)^n \geq a_1a_2\ldots a_{n-1}\left(\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}\right)$$
$$\left(\frac{a_1+a_2+\ldots+a_{n-1}}{n-1}\right)^{n-1} \geq a_1a_2\ldots a_{n-1}$$
$$\frac{a_1+a_2+\ldots+a_{n-1}}{n-1} \geq \sqrt[n-1]{a_1a_2\ldots a_{n-1}}$$

Therefore, P(n) implies P(n−1). To summarise, P(1) is true, and for any n, P(n) implies P(2n), and P(n) implies P(n−1). Therefore, P(n) is true for all natural n. This completes the proof.

Two-dimensional induction

This is the last form of induction which could be expected to appear until well into a university maths degree. Here, we switch from a proposition in one natural variable P(n) to one in two, P(m,n). However, there are actually various methods of completing the induction step:

1. Show P(1,1) as a base case. Then, show that for any m and n, P(m,n) implies P(m+1,n) and P(m,n) implies P(m,n+1).
2. Show P(a,1) for any a and P(1,b) for any b as base cases. Then, show that for any m and n, P(m,n) implies P(m+1,n+1).
3. Show P(a,1) for any a and P(1,b) for any b as base cases. Then, show that for any m and n, P(m+1,n) and P(m,n+1) together imply P(m+1,n+1).
4. Let Q(k) be the proposition that P(m,n) is true for all m+n=k. Then, perform normal induction on Q(k).

The first example will be a relatively well-known combinatorics result which is not normally proved by induction. The problem is as follows:

Show that for non-negative integer variables $x_1, x_2, ..., x_m$ and integer n, the number of different combinations such that $x_1 + x_2 + ... + x_m = n$ is equal to $\binom{n+m-1}{n}$

Here $\binom{n+m-1}{n}$ is the "choose" function, whose values appear in Pascal's triange and binomial coefficients. To prove this statement, we choose method 3, and start with the base cases: If n=1, then there are clearly only m combinations as one variable must be one and the rest must be zero. Accordingly, $\binom{1+m-1}{1} = \binom{m}{1} = 1$.

On the other hand, if m = 1, then there is clearly only one solution to the equation $x_1 = n$, and $\binom{n+1-1}{n} = \binom{n}{n} = 1$.

For the inductive step, we split into two cases: $x_1 = 0$ or $x_1 \geq 1$. In the first case, x_1 can be ignored and there are m-1 variables on the left-hand side remaining. This is equivalent to the case P(m–1,n). In the second case, take 1 away from both sides to get:

$$(x_1 - 1) + x_2 + \ldots + x_m = n - 1$$

But since x≥1, x_1-1 is simply non-negative and is equivalent to the other variables, and this case is equivalent to P(m,n–1). The total number of combinations is just the sum of combinations within these two cases, and by considering the properties of Pascal's triangle, we see that

$$\binom{n+(m-1)-1}{n} + \binom{(n-1)+m-1}{(n-1)} = \binom{n+m-2}{n} + \binom{n+m-2}{n-1} = \binom{n+m-1}{n}$$

This completes the inductive step, and the proof. As an exercise, try this question, using method 1) outlined at the start of this section:

Let f(m,n) be a function on natural numbers m and n such that f(1,1) = 1 and:

$$f(m+1,n) = f(m,n) + 2(m+n)$$
$$f(m,n+1) = f(m,n) + 2(m+n-1)$$

Show that $f(m,n) = (m+n)^2 - (m+n) - 2n + 2$.

Advanced methods

As a rigorous explanation of these methods would be well beyond the scope of this article, I will mention them briefly along with a few common applications of each. One thing that surprised me in the process of writing this article was the volume of theorems in advanced set theory and group theory that used induction, relative to those in more accessible areas. Therefore, I will attempt to give a glimpse of some notable results and leave further research to those who are particularly interested.

Structural Induction
Structural induction is a technique common in graph theory and computer science concerning recursively defined structures such as formulas or tree graphs. Essentially, for some structure x, a proof of a property P(x) constitutes showing that P holds on some minimal substructure such as a single node of a graph. Then, it is shown that if P holds for the immediate substructures of some S, then it must also hold for S.

The most important theorem that uses structural induction in its proof is Łoś's theorem, also known as the fundamental theorem of ultraproducts (which are defined as a quotient of a direct product of a family of structures).

Transfinite Induction
Transfinite induction is induction on ordinal numbers, which is a generalisation of the natural numbers to counting with infinite sets. The process is the same except that an extra step needs to be performed to jump forward to each limit ordinal, which is an ordinal number that cannot be reached by incrementing from smaller ordinal numbers.

One surprising example of transfinite induction is its use in proving Goodstein's theorem, which is a theorem about recursive sequences called Goodstein sequences. First, start with any natural number as the first term. Then, write the number as a sum of powers of 2, and write each exponent as a sum of powers of 2, etc. (for example, $19=2^{2^2} + 2 + 1$). This is called the hereditary base 2 representation. Then, replace each 2 in the expression with a 3, and subtract 1. This is the second term of the sequence. The nth term is found by writing the previous term in hereditary base n representation, then replacing each n with n+1 and subtracting 1. The statement of the theorem is that every Goodstein sequence must eventually reach 0.

The proof of Goodstein's theorem roughly entails showing that it is possible to construct a sequence of ordinal numbers parallel to the Goodstein sequence which is strictly decreasing and,as such, terminates. Therefore, the Goodstein sequence must also terminate at 0. This result is historically significant since it is one of the first true statements to be shown to be unprovable in a system of logic called Peano arithmetic, as was proved by the mathematicians Laurie Kirby and Jeff Paris in 1982.

Fermat's Last Theorem
The details of Andrew Wiles' famous proof of Fermat's Last theorem are of course inaccessible to the common person, but I feel it would be remiss to not mention it briefly. Fermat's Last Theorem is named so because it was the last theorem left unproved by the great Pierre de Fermat for hundreds of years after his death. There is a well-known story that Fermat wrote the theorem in a margin of his textbook, along with the words "I have a marvellous proof of this theorem, which this margin is too small to contain." Nowadays, mathematicians doubt whether he did indeed find a valid proof with the mathematical tools available to him at the time. In any case, the theorem reads:

For any integer n > 2, there are no positive integers a, b, and c such that $a^n + b^n = c^n$.

Wiles' proof, very briefly, works by rearranging the above equation to that of an elliptic curve, which is a family of curves typically described by the equation $y^2 = x^3 + ax + by$. Then, he proved a special case of the Taniyama-Shimura conjecture, otherwise known as the modularity theorem, which establishes a link between those elliptic curves and certain types of analytic functions called modular forms. Using this link, he showed that the elliptic curve resulting from Fermat's Last theorem failed to satisfy certain properties when turned into a modular form, therefore showing that the original equation has no solutions. As it happens, induction is used multiple times in intermediate steps of the proof, giving it the status of perhaps the most important applications of induction in recent years.

Conclusion

Mathematical induction is a unique method of proof in mathematics with highly varied applications requiring insight and precise reasoning. I hope that through the techniques I have presented in this article, the reader may soon spot more and more inductive reasoning in their day to day maths, or at least gain an appreciation of how it works and how it has contributed to the subject as a whole.

Miscellaneous notes

For anyone familiar with these notations, the principle (or "axiom") of induction can be written in second-order logic and first-order ZFC set theory respectively as follows:

$$\forall P(P(0) \land \forall k(P(k) \to P(k+1)) \to \forall n(P(n)))$$
$$\forall A(0 \in A \land \forall k(k \in A \to (k+1) \in A) \to \mathbb{N} \subseteq A)$$