37th IAA SYMPOSIUM ON SPACE POLICY, REGULATIONS AND ECONOMICS (E3)
Assuring a Safe, Secure and Sustainable Environment for Space Activities (4)

Author: Dr. Nathaniel Dailey
The MITRE Corporation, United States, ndailey@mitre.org
Dr. Thomas Groesbeck
The MITRE Corporation, United States, tgroesbeck@mitre.org
Ms. Karen Lee
The MITRE Corporation, United States, kflee@mitre.org
Mr. Eric Sundby
University of North Dakota, United States, lang.sundby@und.edu
Ms. Zhanna Malekos Smith
The MITRE Corporation, United States, zmalekossmith@mitre.org
Dr. Ruth Stilwell
Aerospace Policy Solutions, LLC, United States, office@aerospacepolicysolutions.com

BEHAVIORAL ECONOMICS IN SPACE: STEERING THE FUTURE OF SPACE GOVERNANCE FOR SECURITY AND SUSTAINABILITY

Abstract

This white paper outlines an innovative strategy for promoting a secure, stable, and sustainable environment for space activities as humanity expands its reach to the Moon and other celestial bodies.
Drawing from the influential work of behavioral economists Richard Thaler and Daniel Kahneman, the paper advocates for the strategic implementation of behavioral economics to address the evolving challenges in space governance arising from the increasing involvement of commercial space companies and the shift in launch operations from Earth to extraterrestrial locales.

Incorporating behavioral economics into space governance frameworks can reshape decision-making processes, bolster international collaboration, and encourage enduring, sustainable practices amidst the complex dynamics of an expanding space industry. The white paper details how an intricate understanding of decision-making biases and heuristics, central to Kahneman's findings, could incentivize safer and more sustainable operations by space entities. For instance, policymakers could exploit biases such as the "status quo heuristic"[1] by setting initial standards for responsible behavior, which organizations are likely to follow, such as transparent sharing of space track data or decommissioning inoperative satellites. Additionally, as "Availability Heuristics"[2] shows, people assess the probability of risks based on how easily examples come to mind. Widely publicized incidents of space debris causing damage can nudge organizations to prioritize the development of debris mitigation technologies. Another potential heuristic known as the "Framing Effect,"[3] presenting the long-term sustainability of space activities as a cost-saving measure for future operations, might encourage companies to adopt such measures now.

[1] The *status quo heuristic* refers to people's tendency to stick with the default option rather than actively choosing alternatives, often due to convenience, familiarity, or perceived difficulty in making changes. A common example of this is seen in workplace retirement plans, where employees often remain in the default investment options provided by their employer, even if alternative plans may offer better long-term benefits. Similarly, smartphone settings are typically left unchanged unless users are prompted or "nudged" to adjust them, such as with a notification encouraging them to enable privacy features or update security settings. This concept demonstrates how small changes in choice architecture can significantly influence behavior without restricting individual freedom of choice.
[2] The *availability heuristic* refers to how people assess the likelihood or risk of an event based on how easily examples of that event come to mind. For instance, after hearing about a plane crash on the news, people may overestimate the danger of flying because the event is fresh and vivid in their memory, even though flying remains statistically very safe. Similarly, widely publicized incidents of space debris collisions might lead space operators to prioritize debris mitigation efforts, as the risk seems more immediate and pressing due to recent high-profile examples. This heuristic shows how recent or memorable events can shape perceptions and influence decision-making, even when actual probabilities remain unchanged.
[3] The *framing effect* occurs when people's decisions are influenced by how information is presented rather than the actual content of the information. For example, a medical treatment with a "90% survival rate" sounds more reassuring than one with a "10% mortality rate," even though both describe the same outcome. In the context of space governance, framing sustainability measures as "cost-saving strategies for future operations" might encourage companies to adopt them, whereas presenting the same measures as "additional upfront expenses" could deter them. This effect highlights how the way options are framed can significantly impact behavior and decision-making, even when the underlying facts are identical.

We propose Thaler's 'nudge theory'[4] to enhance space operations efficacy and safety, suggesting that subtle incentives can significantly improve spaceflight safety and prevent conflicts. Applying principles of behavioral economics to space situational awareness, we can foster an environment where information exchange and collaborative efforts are the norm, thereby contributing to an increased collective understanding and stewardship of space.

Strategically employed behavioral insights can lead to renewed approaches that influence perceptions and actions in space, leading to a comprehensive approach encourages positive behavior among spacefaring nations and entities and indirectly deters hostile or negligent activities. This inclusive approach maintains a spirit of cooperation among nations and secures the integrity and sustainability of space for future exploration. The white paper concludes by contemplating the profound impact of incorporating these behavioral economics principles within the space sector, underscoring their potential to bolster international partnerships and inform the development of global norms and treaties that ensure the sustainable use of space.

[4] *Nudge theory*, popularized by behavioral economist Richard Thaler, suggests that small, subtle interventions can influence people's behavior without restricting their freedom of choice. For example, placing healthier foods at eye level in grocery stores "nudges" consumers to choose them over less healthy options, even though both are available. In space governance, nudges might include offering incentives for companies to share space situational awareness data or providing recognition for sustainable practices, gently steering them toward more responsible behaviors without imposing strict regulations. The theory shows how thoughtfully designed environments can promote better decision-making by harnessing natural human tendencies.

Introduction to Economics and Behavioral Economics

Category	Examples	Classical Economics	Behavioral Economics
Investing in the public interest	Developing new technologies, cooperating on debris mitigation and removal	Actors care only about their outcomes. Governments must impose taxes or subsidies to incentivize prosocial behavior.	Actors care about the outcomes of others (envy, altruism, fairness, etc.) Good behavior can be incentivized with nonmonetary rewards. Actors may prefer actions that align with their worldview and self-image.
Investing in the future	Adopting shared standards, responsible end-of-life operations	Actors have time-consistent preferences and accurately predict future goals: good investments occur immediately unless conditions are expected to change	Actors may prefer to procrastinate on long-term investments; this can be exploited by asking for near-term commitments rather than immediate action. Future focus can be manipulated by priming.
Investing in information	Space situational awareness, data sharing agreements	Actors correctly assess risks and pay for additional information based on an expected increase in utility.	Actors may have beliefs that lead them to under-invest in data. Actors may make choices based on the behavior of others rather than their own observations.

Why Behavioral for Space?

1. Externality-heavy environment

First, the orbital environment is exceptionally vulnerable to externalities[5], the effects of one decisionmaker on everyone else. For example, a collision between two satellites could create debris that could threaten other spacecraft for decades. A higher degree of cooperation may be required in such an environment.

2. Diverse kinds of independent actors, no single sovereign authority.

Second, there are many different states and non-state actors active in space. This diversity of actors precludes the classical solution to externalities, namely having a single benevolent government use taxes, subsidies, and regulations to enforce prosocial behavior. Behavioral economics provides a richer understanding of how to encourage responsible use of shared resources among peers and volunteers.

Space is a common pool resource with an emerging polycentric framework of governance. While the Outer Space Treaty places a burden on each state for the continual supervision of objects launched from its territory, it does not provide a framework for the interaction between sovereign authorities or a hierarchy for managing space activities. For this reason, some have argued that space is ungovernable. However, Ostrom's theory on common pool resources illustrates the capability and the desire of users to consider sustainability[6] in determining their behavior, allowing for the development of cooperative

[5] *Externalities* refer to the unintended side effects of an individual's or organization's actions that affect others, often without being reflected in costs or benefits to the responsible party. In space governance, a major externality is space debris. When one satellite operator leaves debris in orbit, it increases the risk of collisions for all other space actors, creating a shared problem without direct consequences for the original source. Addressing these externalities requires cooperative international efforts, such as debris removal policies or shared investment in collision avoidance technologies, to ensure the sustainability of space as a common resource.

[6] A single satellite collision can have far-reaching effects on Earth, disrupting vital services such as GPS, telecommunications, and air traffic management. For example, the 2009 collision between an inactive Russian satellite and an active Iridium communications satellite created thousands of debris fragments that continue to pose risks to operational satellites today. Without global regulations to enforce responsible space behavior, such collisions could become more frequent, threatening the infrastructure we rely on daily. Behavioral economics offers a path to mitigate these risks by encouraging space actors to adopt safer practices. For instance,

models to emerge. This concept provides an opportunity to explore how efforts to coordinate and share information can lead to more sustainable space behaviors from a global perspective. (Teresa Hutchins)

3. New industry, lack of established norms, unknown unknowns (cognitive inertia, benchmarking)

Third, the current state of the space domain is an exotic environment with a relatively short history, so it has few established norms, and individuals may have trouble conceptualizing its risks and opportunities or finding applicable norms and mental models from other domains. In a decision-making environment with limited historical data, decisions may be more sensitive to priming, herding, the availability heuristic, and other behavioral effects.

4. Aspirational & futuristic

Finally, space is an inherently aspirational and symbolic domain that countries closely associate with their technological achievements, contribution to human knowledge, future prosperity, and status in the international community. As we move from dollars for consumer staples to these more lofty and abstract objectives, it becomes necessary to explore how the psychology of self-image, belonging, and reciprocity should be modeled.

NOTES FOR ECONOMIC EXERCISES

Since the 19th century, the classical approach to economic theory has been defined by mathematical optimization, namely, for {1,2,3 ...} we find the action which maximizes their happiness (utility function) (, ...). Note that the utility function is a mathematical convenience that is not measurable in its absolute value; instead, researchers using historical data or designing experiments infer *relative* values of (, ...) using "revealed preferences." If person buys an apple instead of an orange when both are priced at $1, we can infer that ( ) > ( ) > ($1).

For illustrative purposes, we often imagine how a benevolent, all-knowing, all-powerful government would arrange the world to maximize total happiness. This "social planner"[7] is represented by ( ) , typically resulting in a higher average utility at the cost of total impracticality; we then consider more feasible policies that would approach this value.

We also consider game theory, which reflects situations where the rewards by each person depend on the actions of other people: (, ... , ...). When each agent maximizes their utility, assuming each other agent also chooses their own utility-maximizing action, this is termed a Nash Equilibrium.

For the next section, we describe behavioral effects in this "utility optimization problem" form.

EXERCISE: PROSOCIAL BEHAVIOR[8]

nudging companies to invest in collision avoidance technologies or framing long-term debris mitigation as a cost-saving measure can help prevent these incidents, protecting the critical systems that underpin modern society.

[7] The *social planner* concept in economics refers to an idealized decision-maker or governing body that seeks to maximize the overall well-being of society by making choices that benefit everyone. For example, a social planner might impose taxes on pollution to reduce environmental damage, aiming to balance individual freedom with societal welfare. In space governance, a social planner might set global regulations for satellite deorbiting or space debris removal to ensure that all space actors contribute to a safe and sustainable orbital environment. This concept represents the idea of centralized, optimal decision-making for the collective good, even if individual actors might prefer different, self-serving choices.

[8] *Prosocial behavior* refers to actions intended to benefit others or society as a whole, such as cooperation, altruism, or helping. In the context of space governance, prosocial behavior might include voluntary data sharing on space situational awareness, investing in debris mitigation technologies, or adhering to international space sustainability guidelines. Encouraging prosocial actions among space actors can lead to collective benefits, such as reducing

Suppose that we have identical actors; each individual has the choice of whether or not to take an action ($\{0,1\}$). Suppose each player's income is

$$y_i = y_0 - a_i + v\frac{\sum_j a_j}{n}$$

meaning that the action is costly but benefits each and every member of the group.

We will initially assume that the utility of each individual is = . Simply summing across all individuals, we have

$$\sum_i u_i = \sum_i y_i = ny_0 - \sum_i a_i + nv\frac{\sum_j a_j}{n} = ny_0 + (v-1)\sum_i a_i$$

If an all-powerful being wanted to maximize the happiness of all actors and > 1, this "social planner" would force everyone to take action ($= 1$) and achieve the overall utility = + (1) .

On the contrary, suppose that every player makes their own choices; without loss of generality, consider player 1. Let us temporarily assume that the other actors have made their decisions already, and let = .

Player 1 decides to participate in the prosocial behavior if and only if

$$y_0 - 1 + v\frac{(m+1)}{n} \geq y_0 + v\frac{m}{n}$$

Or equivalently, . If there are too many actors or the potential value of the shared project is too small, then player 1 will choose $= 0$, and, following the same logic, all of the other players choose 0, achieving the net benefit = . Thus, we have a loss from our lack of coordination whenever $1 <$ $<$.

Relative Income, Spite

In classical economics, we tend to assume that participants in a market only care about their own outcomes. Behavioral economics has found that study participants often tend to observe the outcomes received by their peers. This can be judged as envy, an instinctive desire for market control, or simply using the experience of others as a baseline and assuming that a similar outcome is the expected payoff.

To assess the effects of such preferences, let the utility of each consumer be as follows:

$u_i = y_i - \alpha\bar{y}$ where $\bar{y} = \frac{\sum_j y_j}{n}$ and $\alpha \in [0,1]$. $\alpha = 0$ gives the case we saw before, while $= 1$ gives an extreme case where each actor only cares about how many dollars ahead of the average actor they are.

As in the previous exercise, we start by considering a Social Planner who wants to maximize the joint utility of all market players:

$$\sum_i u_i = \sum_i y_i - n\alpha\bar{y} = (1-\alpha)\left(ny_0 + (v-1)\sum_i a_i\right)$$

If < 1 and > 1, then, as before, joint utility is maximized by all players being ordered to take the prosocial action. If $= 1$, then, interestingly enough, any action results in the same utility of 0. The utility for each individual is

$$y_i - \alpha\bar{y} = y_0 - a_i + v\frac{\sum_j a_j}{n} - \alpha\frac{\sum_i y_i}{n}$$

Now, without loss of generality, consider the decision of Agent 1. As we saw previously, increasing from 0 to 1 increases by $-$ 1 units and increases by (1). Therefore, the individual actor chooses the action if and only if

$$-1 + \frac{v}{n} - \alpha(v-1) \geq 0$$

collision risks and preserving a safe space environment for all, even though individual actors may bear some upfront costs.

Effective policies can nudge organizations toward these behaviors by highlighting shared long-term gains for the space community.

Or, equivalently, if (1 + (1)). Or $\frac{(\quad)}{(\quad)}$

This presents a larger hurdle to cooperation than the previous problem: as approaches $1/$ the value of necessary to ensure participation grows to infinity, and for $> 1/$ participation becomes impossible. As before, a more valuable project and less participants make participation more sustainable when decisions are made by individual agents.

Intuitively, the problem here is that the envy factor increases the cost of the prosocial behavior: the agent "falls behind" by paying 1, then "falls behind" again when all of the other participants reap the rewards of their hard work. By comparing to the Social Planner's problem, we can also view as an "aversion to being a sucker": while joint utility is maximized.

How should policymakers respond to this kind of hypercompetitive environment? Suppose we offer a reward for participating in the prosocial activity. The payoff to the player becomes:

$$u_i = y_i - \alpha\bar{y} + xa_i = y_0 - a_i(1-x) + v\frac{\sum_j a_j}{n} - \alpha\frac{\sum_i y_i}{n}$$

And participation occurs if and only if

$$x - 1 + \frac{v}{n} - \alpha(v-1) \geq 0 \quad \rightarrow \quad v\left(\frac{1}{n} - \alpha\right) \geq 1 - x - \alpha$$

Thus, as before, needs to be lower than $-$, but a wider range of values can support participation in the prosocial activity.

Note that we could also set the utility function to be:

$$u_i = y_i - \alpha\bar{y} + xa_i - \tilde{\alpha}\bar{x}$$
$$= y_0 - a_i(1-x) + v\frac{\sum_j a_j}{n} - \alpha\frac{\sum_i y_i}{n} - \tilde{\alpha}\frac{m}{n}$$

So, the actor takes some measure of pride in having their award for prosocial behavior, *especially* when others do not. This effect could be especially valuable in a modified scenario where the group reward function rather than ——, is a convex function where incentivizing participation is most difficult when the adoption rate is lowest.

Altruism and Costly Signaling

Numerous laboratory studies have found that participants favor actions that help their peers, even at a (mild) cost to their own outcomes, reflecting generosity or altruism.

Amartya Sen provided an

> "Where is the railway station?" he asks me. "There," I say, pointing at the post office, "and would you please post this letter for me on the way?" "Yes," he says, determined to open the envelope and check whether it contains something valuable.

Numerous psychological and evolutionary explanations have been given for this behavior:

1. An internal sense of being rewarded by being "fair" or even "generous."
2. Increased social status from being observed to be "generous" and, moreover, having the extra resources necessary to be generous (the handicap principle)
3. The assumption is that offering another person a favor will make them offer you an unspecified favor at some point in the future.
4. Evolutionary kin selection, i.e., helping someone related to you survive, prosper, and raise healthy children, will help you pass on *some* of your genes

We can model this preference by using our utility from the previous section while changing the sign of :

$$u_i = y_i + \alpha\bar{y} = y_0 - a_i + v\frac{\sum_j a_j}{n} + \alpha\frac{\sum_i y_i}{n}, \alpha \geq 0$$

A Social Planner hoping to maximize the happiness of all acters will maximize this equation:

$$\sum_i u_i = \sum_i y_i + n\alpha\bar{y} = (1+\alpha)\left(ny_0 + (v-1)\sum_i a_i\right)$$

Unlike the envy model, this gives an unambiguous result that the optimal social outcome is $= 1$ for all .

When participation is left up to the individual, participation now requires that:

$$-1 + \frac{v}{n} + \alpha(v-1) \geq 0 \quad \rightarrow \quad v\left(\frac{1}{n} + \alpha\right) \geq 1 + \alpha$$

Which is clearly more feasible for a wider variety of values than

A considerable literature has explored how to increase altruism in a community. Emphasizing shared values and characteristics across a group has been emphasized. Priming individuals to feel well-resourced (e.g., after a recent raise) and secure (e.g., after being asked to imagine what they would do if they were physically indestructible) seems to increase altruistic behavior.

Utilizing Altruism: Signaling and Self-Image

Suppose that altruism varies throughout a community from high ( ) to low ( ). Altruistic behavior could be encouraged by promising to announce and publicize the names of participants. Investors with future business opportunities might seek an -type partner who will consider their counterparty's profits alongside their own. $= 1$ could be used to identify , and provided that -type actors still find it in their best interest to pick $= 0$ even with these business opportunities, then some intermediate actors $<$ $<$ who may have abstained from the prosocial behavior on altruism grounds alone may decide to participate. *Note: this could be formalized out with some equations.*

A more psychological approach highlighted by Ayres, Raseman, and Shih (2010) is to focus marketing for the prosocial behavior on statistics regarding the actor's peers. Most people prefer to think of themselves as prosocial individuals ( ), so showing a $= 0$ individual that is close to (most of their peers are already participating in the prosocial project) shows that they are either at the selfish end of the distribution of types across their community ( ) or they have miscalculated their own utility function, perhaps by underestimating the value of the project to the community . As the latter conclusion is more palatable, they may reorient to $= 1$.

Behavioral Economics and Norm Developments

TG: An equilibrium where everyone invests in prosocial behavior due to signaling, internalizing altruism, or internalizing the value of our "achievement award" can be viewed as a mathematical model for the development of a new norm. Who would create the norm, champion it, and be targeted for early adoption, and what does the mystery project in this exercise represent?

Exercise: Long-term Investments

Suppose that a company is considering a project that will cost $5 to implement[9] and immediately provide $1 once per year for the rest of time while also providing some benefit to the rest of the world so that policymakers want to incentivize as many actors as possible to implement the project.

A classical Samuelson method is to model the actor as maximizing the equation. = + + + = ; this is also often called the "Net Present Value" in business schools.

If our decisionmaker has = 0.9, the net utility for the project we are considering is thus:

$$\underbrace{-5+1}_{t=0}+\underbrace{.9(1)}_{t=1}+\underbrace{.9^2(1)}_{t=2}+\cdots=-5+\frac{1}{1-.9}=5>0$$

Thus, they chose to undertake the project immediately. Note that delaying the project is not desirable: under these "time-consistent preferences," "anything worth doing is worth doing immediately."

$$\underbrace{0}_{t=0}+\underbrace{0.9(-5+1)}_{t=1}+\underbrace{.9^2(1)}_{t=2}+\underbrace{.9^3(1)}_{t=3}+\cdots=-5(0.9)+\frac{0.9}{1-.9}=4.5<5$$

Hyperbolic Discounting[10]

However, numerous studies have found that delaying benefits from the current period to the next period is far more distasteful than delaying from, say, year 3 to year 4. In layman's terms, the agent believes they are *generally* patient and forward-looking but happen to be "in a tight spot at the moment where they have to be focused on the here and now."

Quasi-hyperbolic[11] preferences can be used to model this discrepancy: suppose instead that in each period decisionmaker tries to maximize the equation = + + + = + . The key here is that there is a steep drop-off in the first period and a slower exponential decay afterward.

Suppose that this particular decisionmaker has values = 0.5 and = 0.9.

At = 0 the value of undertaking the project immediately is:

$$\underbrace{-5+1}_{t=0}+\underbrace{(0.5)0.9(1)}_{t=1}+\underbrace{(0.5)0.9^2(1)}_{t=2}+\cdots=-5+1+\frac{0.9*0.5*1}{1-.9}=0.5$$

The decisionmaker will, at this point, realize that they may enjoy a higher utility if they do nothing at t=0 and undertake the project at t=1, with the net payoff:

$$\underbrace{0.5(0.9)(-5)+(0.5)0.9(1)}_{t=1}+\underbrace{(0.5)0.9^2(1)}_{t=2}+\cdots=-2.25+\frac{0.9*0.5*1}{1-.9}=2.25\ (in\ t=0\ dollars)$$

However, if they do this, then at t=1 they will face an identical choice.

The benefit at = 1 of executing the project immediately is:

$$\underbrace{-5+1}_{t=1}+\underbrace{(0.5)0.9(1)}_{t=2}+\underbrace{(0.5)0.9^2(1)}_{t=3}+\cdots=-5+1+\frac{0.9*0.5*1}{1-.9}=0.5\ (in\ t=1\ dollars)$$

---

9 For the sake of simplicity we use small round numbers in this example; in the space economy setting you could imagine these values in hundreds of millions to billions of dollars.

10 *Hyperbolic discounting* refers to the tendency for people to prefer smaller, immediate rewards over larger, delayed ones, even when waiting would yield a better outcome. For instance, someone might choose to spend $100 today rather than save it and receive $150 in a year, even though the latter option is more financially beneficial. In the context of space sustainability, hyperbolic discounting could explain why space operators might delay investing in long-term debris mitigation strategies, opting instead for short-term cost savings, despite knowing that future cleanup costs could be much higher. This concept highlights how people tend to undervalue long-term benefits, leading to procrastination or shortsighted decision-making.

11 A *true* hyperbolic preference would follow the equation = —; the Quasi-Hyperbolic preference formula is an approximation of this formula.

While from the $= 1$ perspective, the benefit of delaying one additional period is:

$$\underbrace{0.5(0.9)(-5) + (0.5)0.9(1)}_{t=2} + \underbrace{(0.5)0.9^2(1)}_{t=3} + \cdots$$
$$= -2.25 + \frac{0.9 * 0.5 * 1}{1 - .9} = 2.25\ (in\ t$$
$$= 1\ dollars)$$

Thus, at each year , the decisionmaker will plan to execute the project at $+ 1$; by induction, we get the paradoxical result that the benefits of the project always outweigh the costs, but the decisionmaker will *never* actually execute the project. While usually applied to individual decision-making, it could be argued that this model helps us understand how much larger policy decisions, such as the US adoption of the metric system and the abolition of Daylight Savings Time, have similarly been trapped in a limbo where it is well understood that their benefits clearly outweigh their costs but they have been "planned for the near future" for a century.

First, if the decisionmaker is aware of this propensity to procrastinate and has no ability to commit themselves to future action, then they may simply undertake the project at $= 0$. This requires the decisionmaker to be "sophisticated" and set up stricter requirements for adoption of the project: if the project's cost is 6, for example, an individual with $= 0.5$, $= 0.9$ preferences might *only* agree to the project if it can be delayed to the next period (see below).

$$Start\ project\ at\ t = 0: \quad \underbrace{-6 + 1}_{t=0} + \underbrace{(0.5)0.9(1)}_{t=1} + \underbrace{(0.5)0.9^2(1)}_{t=2} + \cdots = -6 + 1 + \frac{0.9*0.5*1}{1-.9} = -0.5 < 0$$

$$Plan\ to\ start\ at\ t = 1: \quad \underbrace{0.5(0.9)(-6) + (0.5)0.9(1)}_{t=1} + \underbrace{(0.5)0.9^2(1)}_{t=2} + \cdots = -2.7 + \frac{0.9*0.5*1}{1-.9}$$

Second, they could be convinced to accept a "commitment device". Suppose at $= 0$ they can apply for a grant and receive a grant of 2 immediately for implementing the project, which must be refunded if the project is not completed by the end of $= 1$.

From the perspective of the agent at t=0, the utility of accepting the grant and starting the project immediately is

$$2 - 5 + 1 + (0.5)0.9(1) + (0.5)0.9^2(1) + \cdots$$
$$= -5 + 1 + \frac{0.9 * 0.5 * 1}{1 - .9} = 2.5$$

While the payoff for accepting the grant and delaying the project for one turn is

$$2 + 0.5(0.9)(-5) + (0.5)0.9(1) + (0.5)0.9^2(1) + \cdots$$
$$= 2 - 2.25 + \frac{0.9 * 0.5 * 1}{1 - .9} = 4.25$$

Note that both of these options have a higher utility than *refusing* the grant or accepting the grant at a later date.

Once we are at $= 1$ the utility of undertaking the project immediately becomes

$$-5 + 1 + (0.5)0.9(1) + (0.5)0.9^2(1) + \cdots$$
$$= -5 + 1 + \frac{0.9 * 0.5 * 1}{1 - .9} = 0.5$$

While the utility of delaying the project one more turn is

$$-2 + 0.5(0.9)(-5) + (0.5)0.9(1) + (0.5)0.9^2(1) + \cdots$$
$$= -2.25 + \frac{0.9 * 0.5 * 1}{1 - .9} = 0.25$$

Thus the grant and its deadline take advantage of the agent's tendency to place an exceptionally high value on present profits (and inability to realize that it they will still have the same present bias in the next period) to make them volunteer for a binding contract which will then force them

to move forward with the project at $= 1$. Note that simply offering a 2 unit subsidy for the

$= 1.8 > 0$

project whenever it gets made would still result in the agent preferring to delay the project indefinitely into the future:

$$(Immediately)\ (-5 + 2) + 1 + (0.5)0.9(1) + (0.5)0.9^2(1)$$
$$+ \cdots = -5 + 1 + \frac{0.9 * 0.5 * 1}{1 - .9} = 2.5$$

$$(delay\ 1\ turn)\ 0.5(0.9)(-5 + 2) + (0.5)0.9(1) + (0.5)0.9^2(1)$$
$$+ \cdots = -1.35 + \frac{0.9 * 0.5 * 1}{1 - .9} = 3.15$$

In the context of the sustainable space environment, many interventions may be seen as a sharp up-front cost for launch providers, satellite manufacturers, satellite operators, governments, etc., whose costs are only justified in the long term. Even if the direct benefit to the decisionmaker will be profitable and rational from the perspective of classical economics, behavioral economics suggests that actors may have hyperbolic preferences that cause them to repeatedly delay the project so that it always remains one year ahead.

The solution proposed here is to have actors volunteer to commit to implementing the sustainability project before a set deadline, with the following strictures:

1. There must be a clear up-front benefit to the commitment, such as positive publicity, influencing new standards, or a substantial pecuniary grant
2. There must be a hard deadline for executing the project with no chance of renegotiation, and thus, the deadline must be generous enough to accommodate delays due to legitimate technical issues.
3. There must be consequences for failing to execute the project equal to the original boon, such as negative publicity or being forced to return the grant.

In summary, deciding whether to undertake some long-term investment is a simple now-or-never choice for an agent in a classical model with exponential preferences. In a behavioral model with hyperbolic preferences, the agent says, "Now isn't a good time, but next year I will start investing in the future, " unaware that they will feel the same way next year and continue to procrastinate on the project. To encourage the adoption of a forward-thinking project, we recommend exploiting this bias by offering actors an opportunity to commit to executing the project before a hard deadline in the near future.

MENTAL TIME TRAVEL

Returning to our classical model, where actors attempt to maximize = + + + = , we might ask where comes from. While it can be argued that keeping pace with inflation, market interest rates, the return on investment of a balanced portfolio adjusted for risk, or a government-mandated discount rate for evaluating public works projects can be crucial, many decisions involving new business models and untested technology may rely more heavily on "gut instinct" than formulaic present value discounting.

In the simple one-time investment mentioned previously, the agent invests immediately if and only if the following holds true:

$$\underbrace{-5+1}_{t=0} + \underbrace{\delta(1)}_{t=1} + \underbrace{\delta^2(1)}_{t=2} + \cdots = -5 + \frac{1}{1-\delta} \ge 0$$

Or equivalently, $\delta \ge {}^4/_5$. The more expensive the start-up cost becomes, the more future-focused the actor needs to be in order to adopt the project:

$$-COST + \frac{1}{1-\delta} \ge 0 \quad \rightarrow \quad \delta \ge 1 - \frac{1}{COST}$$

An extensive series of clinical trials, such as the work of Prof. Hal Hershfield, have shown that "mental time travel" can be used to increase people's preference for future well-being . For example, students who had been shown digitally altered photos of themselves 40 years in the future were willing to save 30% more money than students who had looked at unaltered photos.

This behavioral marketing literature suggests that a crucial part of convincing actors to adopt a future-focused attitude is presenting them with vivid imagery representing a possible future and *convincing actors to participate in thought experiments about the future.*

EXERCISE: ENDOWMENT EFFECT, PSYCHOLOGICAL INERTIA, NUDGING

One interesting finding in numerous controlled trials is the endowment effect: people prefer to keep what they have and show a bias against "messing with a good thing" by trading for a new option. This is shown in controlled trials which randomly assign products to study participants: participants who have been gifted a mug, for example, quote a much higher price for the mug than the participants who did not receive the mug. We own things that we value: classical economics says we chose to acquire valuable things, but behavioral economics adds that having something makes us value it more.

Suppose that the actor is trying to decide whether or not to take the action at time , with benefit and cost . Suppose there is also cost for changing their choice relative to their choice from the previous year. Their goal is to maximize the following utility function:

(,) = ( ) | |

Given a discount factor , actors will switch from = 0 to = 1 if and only if:

$$\sum_{s=0}^{\infty} \delta^s (v - c) - d \geq 0$$

$$\frac{(v - c)}{(1 - \delta)} \geq d$$

$$v \geq c + d(1 - \delta)$$

Note that it can be easily shown that we do not need to consider delaying switching or switching back and forth.

An actor who is already playing the action will keep doing so (= = 1) if

$$\sum_{s=0}^{\infty} \delta^s (v - c) \geq -d$$

$$\frac{(v - c)}{(1 - \delta)} \geq -d$$

$$v \geq c - d(1 - \delta)$$

A higher value of makes it more likely actors will maintain their current course of action (=), behavior referred to as *status quo bias*.

In classical economics, reflects *external* penalties, such as time spent writing new contracts, recruiting new employees, sales commissions, etc., often called "adjustment costs". In behavioral economics, we can also include the *internal* penalty of changing behavior, such as the cognitive load of making an "active" decision, the fear of making a poor choice or accepting that your previous choices were poor, broadly termed *psychological inertia*.

This finding has two major implications for promoting good behavior in the space domain:

1. Without changing any incentives, behavior can be promulgated by simply presenting it as the default choice and requiring the actor to *opt out* rather than *opt in*. For example, more people chose to become organ donors when a driver's license form was changed so that applicants need to check a box to *decline* to be a donor. This "nudging" could be applied to the space domain by working with licensing authorities, satellite manufacturers, launch providers, etc. to present participation in space debris management programs as part of the default bundle.

2. Psychological inertia underlines the importance of introducing good norms of behavior at an early stage of development. Working with startups to ensure that collision avoidance, situational awareness, and responsible disposal are incorporated into their first wave of satellites may be easier than approaching them after they have locked in on a mature business model and asking them to change their mission profile or hardware.

Practical Application, Policy & Administrative

TG: We could use some examples of projects for space sustainability that will have good long-term returns for the satellite operators, partner countries, launch providers, etc., which are willing to make a costly change in operations, pay a startup cost to create an international community organization, etc.; where participation is difficult to *mandate* our Exercise provides suggestions for how actors can be convinced to *voluntarily* participate, either by getting them to sign a medium-term pledge which will be costly to back out of once they are forced to "put their money where their mouth is" or by exploiting behavioral marketing techniques to prime decisionmakers for long-term thinking.

The United Nations Office of Outer Space Affairs has explicitly endorsed initiatives promoting space sustainability, but has primarily focused efforts by individual states or international organizations. Mandating norms on an international scale proves difficult for almost any sector, and under current geopolitical realities could prove to be even more challenging. Mandating norms under domestic legislation can also be seen negatively by industries, especially those as young and maturing as space. However, by encouraging norms and implementing novel projects potentially via private entities, the space sector may witness good long-term returns for satellite operators, partner countries, and launch providers. One key example that is highlighted by some industry leaders, such as the European Space Agency and Voyager Space, is the Space Sustainability Rating.[12] The Space Sustainability Rating uses a tiered scoring system that utilizes existing metrics that have been developed and published by government agencies and others to measure sustainable actions in space. Once these metrics are applied to a company or specific mission, the data is verified by the company supplying the rating firm with proper technical documentation related to a capability's application to sustainability. After this process is complete, the corresponding rating is then awarded to the company.[13]

One action that could be taken could include industry associations encouraging companies to voluntarily sign a medium-term pledge that focuses on developing and implementing technological processes that enable sustainable practices in space. Encouraging sustainable methods is intended to create an environment where reversing the practices could prove more costly than simply maintaining them. Including language promoting space sustainability in an industry association's code of ethics may effectively encourage acceptance of norms amongst members. For example, the American Institute of Aeronautics and Astronautics, amongst many other organizations, holds a code of ethics to establish norms in their sector,

[12] Voyager Space, “Why Sustainability in Space is Vitally Important: On Orbit and Earth,” *Voyager Space*, November 1, 2023 https://voyagerspace.com/insights/why-sustainability-in-space-is-vitally-important-on-orbit-and-earth/#:~:text=The%20Space%20Sustainability%20Rating%20provides,of%20organizations'%20debris%20mitigation%20efforts.

[13] *Promoting sustainable behavior of space actors*. Space Sustainability Rating. (n.d.). https://spacesustainabilityrating.org/

including avoiding conflicts of interest and promoting the protection of proprietary information across the industry.[14] Establishing sustainable practices in space as another guideline in the code might serve as a nudge to member companies to be included in the mainstream sector.

Third-party ratings and non-binding codes of ethics are used across multiple industries, as these mechanisms predominantly promote the acceptance of norms from a more social and philosophical standpoint. By following such examples and adopting non-binding methods to advance norms in space sustainability, the sector may avoid overbearing regulations or government mandates that could negatively affect space development.

Exercise: Underinvestment in Information

One common finding is that market participants underinvest in data, relying on naïve trending, biases, and mental models when hard data and better empirical predictors are available.

Each year, let us assume that agents play the following game:

1. Nature determines if an incident occurs this year (= 1) with probability or it does not occur (= 0). Agents to do not observe or p, but their confidence in different possible values of the probability of the incident is represented by the distribution ∼ (,).
2. Each agent decides whether to pay 1 to see a forecast (= 1) or do nothing (= 0). Payers receive early warning of .
3. (= 1) agents who saw (= 1) can take costly action to avoid the incident becoming an *accident*. All other agents suffer an *accident* (= 1) where Pr(|) = is the probability of an unmitigated incident becoming an accident.
4. Agents update their values of , based on and observed (if any) using Bayes' Rule.

It can be shown (*ADD PROOF LATER)* that individuals with low values of , and high initial guesses = —— will decide to purchase the data each period: in the long run they will likely continue to invest in the early warning if and only if it is profitable. On the other hand, if individuals have a low value of = —— with a high value of , representing (erroneously) high confidence in their prediction, they will decide not to purchase the data, and in the absence of regular updates they will only find their mistake through repeated accidents (= 1). In this context, investing in the data serves two purposes: it ensures that the agent is protected from suffering an unmitigated disaster in the current period and allows the agent to see past the "noise" from and more rapidly refine its value of . By contrast, an actor who is overconfident in an overoptimistic risk assessment burrows their head deeper into the sand until it is too late.

A policymaker may be concerned for the welfare of the overconfident optimists (or externalities from their accidents) and has several potential remedies:

1. Provide the data for free. If taxes fund this, it would be equivalent to forcing all actors to select (= 1).

[14] *AIAA code of Ethics*. www. (n.d.). https://www.aiaa.org/about/Governance/Code-of-Ethics

2. Provide free historical data on incidents (, , ...). This would allow non-payers to update , while still incentivizing payment for the data to enjoy the first-order benefit of accident avoidance.
3. Host workshops to help actors discuss and better understand their prior predictions , . Note that the goal might not even be to *increase* the actor's prediction , but rather to decrease , to promote investment in data for experimentation.

One serious concern is how we define "incident." If the definition of an incident is widened while the definition of an accident is kept constant, becomes very small. While data buyers receive more readings of to identify , they are made worse off due to the mitigation actions they keep paying for, which have a probability 1 of proving unnecessary. It may be profit-maximizing in this case for an agent to forego the data altogether to avoid being spooked into unnecessary bills.

Extension: Rational Herding

Suppose many agents are playing the game listed above. Rather than making arguments to attempt to influence some particular agent's beliefs , or their action it may be sufficient to show them the participation rate of their peers simply . An overconfident optimist, presented with the possibility that many of their peers must conclude the following:

1. Their prior beliefs , are unusual, OR their peers have suffered multiple accidents
2. Since investing in the data, their peers have observed many incidents which they have successfully avoided through timely action so that their posterior estimates of justify continued investment in the data.
3. These factors suggest that that is much higher than anticipated, and they should join their peers in investing in the data.

What is interesting here is that it becomes rational for agents to ignore their own observations and focus on the observations that it is *implied* that their peers have collected. This effect can lead to bubbles or panics in some settings as the self-reinforcing beliefs become divorced from actual signals. This effect is somewhat mitigated in this example by the fact that the behavior we are introducing is, itself, an independent source of information. To use a somewhat fanciful example, if a social media hoax about nuclear power led everyone in America to begin carrying a Geiger counter at all times, it would not take long for people to begin realizing that the radiation readings that their unit was picking up did not justify this behavior.

International Development & Capacity Building, Trusted Allies & Partners

TG: New spacecraft operators in the developing world and new space startups may lack the facilities for space situational awareness and dismiss the caution advised by status quo space players like the US government (perhaps in much the same way that the Thirteen Colonies chafed at the UK's efforts to restrict their expansion west of the Appalachians for their safety the good of the international community). Providing free or subsidized live data that is readily interpretable, providing historical data that readily shows the size of collision risks encountered in the past, and hosting talks to decrease overconfidence could do much to ameliorate people's hesitance to "look a gift horse in the mouth" by paying a large amount for data with scary results.

Global capacity-building activities are beneficial because they help foster conditions for states to cooperate on shared goals and can strengthen the formation of a rules-based international order.

States are in the process of exploring how to codify norms of responsible behavior in the space domain. When it comes to capacity building and developing accountability mechanisms in space governance frameworks, multilateral framework solutions are only as effective as the strength of the partnerships and implementation efforts to sustain them.

Drawing a parallel from cyberspace accountability to outer space, according to James A. Lewis, a senior researcher with the Center for Strategic and International Studies, an effective international cyber strategy must focus on the following three elements: "how to build resilience, how to create a collaborative defense, and how to produce accountability in cyberspace[.]"[15]

In terms of agreed-upon governance frameworks, the 1967 UN Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space (OST) is the seminal document. The OST prohibits national claims of sovereignty, establishment of military bases and weapons, or placement of weapons of mass destruction in orbit or on celestial bodies.

Another major treaty is the Liability Convention of 1972. This convention is an agreement for States to take full responsibility for any damages caused by their space objects. However, This is being tested by the rapid evolution of the civil space sector. Overall, the OST, combined with the United Nations Charter and other relevant international laws, represents "the essential framework for the peaceful exploration and use of outer space for the benefit of all nations."[16]

Traditionally, the law is slower in evolving to respond to new technological developments. However, identifying the legal gaps and barriers is necessary to cultivate a robust accountability ecosystem and shape how actors interact in the emerging commercial space economy.

Conclusion

The white paper has laid out a compelling case for integrating behavioral economics into the governance framework for space activities, addressing the significant challenges of ensuring security, stability, and sustainability in this rapidly evolving domain. Drawing on well-established principles from behavioral economics, such

[15] James Andrew Lewis, "Deterrence and Cyber Strategy," CSIS, November 15, 2023, https://www.csis.org/analysis/deterrence-and-cyber-strategy.

[16] European Union, "EU Joint Contribution on the Works of the Open-Ended Working Group on Reducing Space Threats through Norms, Rules and Principles of Responsible Behaviours: Fourth Part: Recommendations on Possible Norms, Rules and Principles of Responsible Behaviour Relating to Threats by States to Space Systems," UNODA, June 15, 2023, https://docs-library.unoda.org/Open-Ended_Working_Group_on_Reducing_Space_Threats_-_(2022)/EU_joint_contribution_to_OEWG_works_on_norms_of_responsible_behaviours.pdf

as nudging, framing, and heuristics, it demonstrates how subtle shifts in decision-making processes can incentivize better behavior among spacefaring entities, fostering international cooperation and encouraging sustainable practices in space.

The paper highlights that traditional governance mechanisms will fall short as we move further into an era where space is no longer solely the domain of nation-states but also includes an increasing number of private entities. Without a central authority to enforce regulations, behavioral insights offer a pragmatic approach to steering space actors toward responsible behavior that recognizes the social and psychological drivers behind decision-making.

Key sections of the white paper emphasize the importance of information-sharing, the dangers of externalities, and the need for long-term investment in space sustainability initiatives. Acknowledging space actors' biases and cognitive barriers, the authors propose actionable strategies, such as utilizing the availability heuristic to highlight space debris risks or leveraging altruistic signaling to promote collaborative efforts among space stakeholders.

The insights drawn from this research are theoretical and have practical implications for policy, regulation, and international collaboration. To that end, the paper provides a roadmap for implementing behavioral economic principles into existing space governance frameworks. It also calls for the urgent need to integrate these strategies into global agreements, fostering a shared responsibility for the sustainable use of space.

We stand on the cusp of a new space era for the global community to adopt innovative approaches to governance that can keep pace with technological advancements. It follows that action should occur now to establish norms, frameworks, and partnerships that secure the long-term viability of space exploration and use. Employing behavioral economics in space governance can create an environment where cooperation, safety, and sustainability become the foundation for all space activities.

In this moment of strategic opportunity, the authors call upon governments, industry leaders, and international bodies to embrace these principles and actively participate in shaping the future of space governance. Our decisions today will determine the safety and sustainability of space for generations to come.

Key Takeaways

This paper introduces the application of behavioral economics to space governance, outlining how simple yet effective strategies can promote safer, more sustainable practices among space actors.

Understanding how decision-making biases shape behavior better equips policymakers and space industry leaders to design interventions that encourage responsible space use, reduce risks from space debris, and foster long-term international cooperation. The following key takeaways summarize the central points:

1. **Behavioral Economics Can Nudge Better Decisions**
   Small, targeted interventions—such as offering incentives for sharing space data or implementing debris mitigation technologies—can influence space operators to adopt safer and more sustainable practices without imposing heavy regulations.

2. **Understanding Human Biases Is Crucial**
   Common decision-making biases, such as the *status quo heuristic* (sticking with default options) or *hyperbolic discounting* (prioritizing immediate rewards over long-term benefits), often lead space actors to delay or avoid important safety measures. By designing policies that account for these biases, stakeholders can guide actors toward better choices.
3. **Framing Matters**
   How a policy or investment is presented can significantly impact its adoption. Presenting space sustainability initiatives as cost-saving in the long run, rather than focusing on the immediate expenses, can encourage companies to take action today.
4. **Proactive Governance Is Essential**
   Without proactive, coordinated governance, the growing risk of space debris and satellite collisions could disrupt critical services on Earth, such as GPS, telecommunications, and weather monitoring. Applying behavioral economics strategies can help mitigate these risks and ensure a safer, more sustainable space environment.

Table 1: Key Takeaways Table

Key Point	Why It Matters	Impact on Space Governance
Behavioral economics nudges better decisions.	Small incentives can steer space actors toward responsible normative behaviors.	Encourages adoption of safety measures like collision avoidance and debris mitigation technologies.
Understanding human biases is crucial.	Biases like *status quo* and *hyperbolic discounting* often lead to poor long-term decisions.	Policies can be designed to counteract these biases, guiding space actors to make better decisions for long-term sustainability.
Framing matters.	The way options are presented influences decision-making.	Presenting sustainability initiatives as cost-saving over time increases the likelihood of adoption, making space safer for all stakeholders.
Proactive governance is essential.	Space is becoming more congested, increasing the risk of collisions that could disrupt vital services.	Applying behavioral economics to governance strategies helps prevent these risks, ensuring the safety and sustainability of space activities for future generations.