CrossMark

CAMBRIDGE
UNIVERSITY PRESS

Hype, skin in the game, and the stability of cooperative science

Adrian Lenardic (iD), Johnny Seales and Anthony Covington

Department of Earth Science, Rice University, Houston, TX 77251-1892, USA
Author for correspondence: Adrian Lenardic, E-mail: ajns@rice.edu

Abstract

We address a recently posed question: 'Why Do So Many Astronomy (and Astrobiology) Discoveries Fail to Live Up to the Hype?' We expand it to cover hype within science in general. Our answer relies on working definitions of hype and skin in the game, as applied to research science, and a game theory model for the stability of cooperative science. Low skin in the game allows internal feedbacks, within the research science community, to initiate increased hype and a drift toward structural instability. The instability leads to the deterioration of cooperative equilibria, which further enhances hype. Along the drift, the number of results hyped as breakthroughs will increase and more claims will fail to live up to the hype. This can lead to the public perception that science is moving backwards and a shift in the perception of what scientists, and science, values. Although a hype instability can be initiated by external nudges, a bigger role is played by the internal dynamics of the system, i.e. the collective of working scientists. Corrections for a drift toward instability should, likewise, focus on internal structure. Proposed external shifts on how research is disseminated will add restrictions to a system that can do more harm than good.

Contents

Introduction

A recent article posed the question 'Why Do So Many Astronomy Discoveries Fail to Live Up to the Hype?' (Falk, 2021). Although phrased as hype in astronomy, many of the specifics discussed relate to astrobiology. The astrobiology connection is timely as the search for life beyond Earth has led to discussions of whether new standards of communication should be put into place for disseminating information related to the search (Green *et al.*, 2021). A goal of such standards would be to avoid overly confident claims that prove to be false over time. This is a worthwhile topic of discussion but it does not get at the cause(s) for overly confident, i.e. hyped, claims to begin with (unless one assumes that the cause is a lack of regulations, a view we do not subscribe to).

The author of the article, that posed the 'why the hype to begin with' question, interviewed a range of scientists and science writers to seek an answer (Falk, 2021). A few points caught our eyes: (1) The

author observed that scientists are hesitant to place the blame on any one part of the process; (2) Interviewed scientists felt it was a problem that could damage science; (3) An interviewee noted that the problem stems from 'everyone having skin in the game'.

Point 1 suggests that the problem is systemic. As such, understanding the structure of the system is critical to addressing the question posed. We would further argue that if the structure is not understood, then any proposed solutions, be they related to new standards of communication or otherwise, may well miss the mark.

Point 2 suggests that the actions of individuals (e.g. a scientist hyping a result) could damage all those that participate in science. Stated another way, the system is one of strategic interdependence: Actions of an actor within the system affect the potential success or failure of the actor and of others within the system. An analysis of the structural problem can thus be approached via game theory.

Point 2 further suggests that a situation that has been stable, a current equilibrium, could lose stability. The equilibrium was not defined beyond calling it 'science'. However, the worries expressed were more specific to the idea that hype undermines cooperative science. This reinforces a game theory perspective as cooperative equilibria have been well studied using game theory models. A game theory perspective requires that the preferences/motivations of actors be as precisely characterized as possible.

Point 3 takes prominence for characterizing motivations. The interviewee who suggested that 'skin in the game' is a motivator for hyping results did not characterize what skin in the science and hype game is. To the best of our knowledge, no one has.

The above sets the motivation for what follows. We start with metrics for hype and skin in game. From there, we argue that skin in the game is not what drives hype. It is a lack of skin in the game. Our analysis will show the degree to which different fields are prone to over-hype. Astronomy and astrobiology fall on the high end but are not unique in that regard. The last aspect of our analysis connects hype to the stability of cooperative science.

What is hype in science?

Our arguments are specific to science and hype. They do not apply to science under hype-free conditions (day-to-day science). To make this precise, we need a working definition of hype. Our definition involves two aspects: dissemination platforms and timing.

A claim is hyped if it is taken beyond its original platform and/or context. Research is normally disseminated in scientific journals, at scientific meetings, and seminars presented to peers. When research is picked up by popular science press, given a media release, presented in blogs, advertised in anyway (on social media platforms and/or interviews), it has moved from a primary to a secondary source platform.

On its own, the above falls short of defining hype. Scientific breakthroughs should be disseminated beyond their original platforms – when Einstein's theory of general relativity was confirmed, by the Eddington lead expedition, it was quite rightly taken from the realm of specialty journals into the public sphere. This suggests that a complete definition requires a temporal element.

Our definition of hype will employ a time to verification/validation metric (τ_v). This can mean different things for a given research result. Results that are in the form of forecasts/predictions have the potential for validation if they are shown to be consistent with future observations. Not all research results provide prediction(s). Many provide explanations for observations. This still allows for a verification time metric. The metric then characterizes reproducibility. That is, how long would it take for another research group to confirm/refute the original result.

Some research studies, modelling studies in particular, are meant to be exploratory and provide insights into the workings of a system. They are not designed to directly address observations and validation, in the traditional sense, cannot be applied. Verification that model equations are being solved correctly can still occur. Added utility checks and community evaluation can also occur (Funtowicz and Ravetz, 1990). As an example, the robustness of model inferences depends on model uncertainties, which are often not fully evaluated at the time a modelling study is presented (the community, that has

interest in the model, can explore it in greater depth after they have been exposed to it). Community assessment can determine if uncertainties degrade model utility (e.g. the use of model inferences/heuristics for decision making and/or hypotheses discrimination). That type of assessment is a means to verify the robustness and utility of model insights. We will group it under τ_v to allow for a general hype metric that can be applied to a range of studies.

We can combine τ_v with the time at which a result is hyped (τ_h), beyond the time it appeared as a primary source (τ_p), to tighten our definition of hype. If $\tau_v > (\tau_h - \tau_p)$, then a result/claim has been hyped (e.g. when a press release appears within days of research first being published). A result is being advertised as significant before it can be verified or confirmed or assessed. Another way of viewing it is that a result is being proclaimed as influential before it can be determined if it will influence future work beyond the time duration over which hype dissipates (i.e. the time when popular science articles, blog posts and/or social media discussions about a result/claim die away). Together with the number of non-primary outlets a result is disseminated on (N_s), this allows us to define a hype factor (H) as

$$H = N_s[\tau_v - (\tau_h - \tau_p)]. \tag{1}$$

If H is zero or negative, a result/claim is not being hyped (that does not mean it is not significant – it just means it has not been advertised as such before verification/validation/assessment can occur). Press releases or other announcements of a research paper that retracts a previous result score zero on the hype scale as, for such cases, the time to refutation is the same as the relative timing of the announcement itself, i.e. $\tau_v = (\tau_h - \tau_p)$. The rise of altmetrics, which quantify the exposure a research paper gets on platforms beyond primary sources, allows hype factors to be calculated with the same relative ease that citation counts can be tracked (Warren *et al.*, 2017; Elmore, 2018).

If a result is advertised as significant before it appears in a primary source publication, then ($\tau_h - \tau_p$) can be negative. This covers scientific claims that are made without having passed through a primary vetting process. In that case, the time to primary publication can become large. As a result, the hype factor can also become large. This is not to say the result/claim could not be verified, validated, or assessed by the community. It only says that, on a scale of hype, it rates high. If the claim is confirmed, refuted, or replicated, then a publication will appear to demonstrate that, and the hype metric will be approximately $2\tau_v$. That type of publication can come from the individual/group that first made the claim or from others. For an exploratory modelling study, assessment can come from the community that is connected to the modelling topic. If none of those occur, then the hype factor tends to infinity. The other way hype can approach infinity is for claims that cannot be refuted, replicated, or tested using the methods of science. This connects hype to a defining factor of scientific claims and, arguably, science itself.

Hype can be initiated via push or pull. Push occurs when scientists contact dissemination outlets to gain a broader audience for their work and/or promote the work via interviews, press clips and/or social media sites. Pull occurs when popular science writers/bloggers contact scientists to write articles/posts about a study and/or when the university/institution scientists work at send out press releases via public relations departments. Once initiated, hype can be extended via a combination of push and pull.

Hype, via push or pull, shares a distinction from hype-free science. In the hype-free process, the arbitrators of determining if a result should be published are peers (experts who serve as referees and editors). The arbitrators determine if sufficient care and rigour has been taken and whether the results will be of value to scientific peers. They do not determine if the results are of interest to the general public nor is there a requirement for validation. The reason being because research papers often put forth new ideas and/or predictions that can motivate the community toward new work (e.g. putting forth a new hypothesis is part of science and it does not require validation before hand – it requires that refutation is possible, often with observations that could be obtained in the future). Completely impartial arbitrators may be an unattainable ideal for humans but efforts are made to maintain impartiality. This becomes difficult in the hype process. For hype via push, the reality is that scientists find their results of value. Even allowing for a high level of perspective from an individual scientist or group in terms of which of their results should

be pushed over others, the process no longer involves impartial arbitrators. For hype via pull, the decision on what is significant can come down to single arbitrators who are not impartial: science writers have content to fill and can be drawn to provocative results that attract more readers; public relations officers are paid to promote results coming from a particular university/institution.

A high hype factor is not necessarily detrimental. Breakthroughs should be hyped. The article that got us to thinking about hype was not about that, however. It was specific to why so many claims are not living up to the hype. It is possible that breakthroughs are decreasing. Even if correct, a rise in the number of hyped results can play a larger role. The relative number of scientific breakthroughs can be constant or increasing, but to the public at large, who will hear principally about hyped results, it can appear that science is moving backwards as more and more claims fail to live up to hype. The question then becomes why an increase in hype. That requires a consideration of motivations and skin in the game.

What is skin in the science and hype game?

The Macmillan Dictionary defines skin in the game as 'being at risk financially because you have invested in something that you want to happen'. From Wikipedia: 'To have skin in the game is to have incurred risk by being involved in achieving a goal'. An author who has written about skin in the game defines it as 'A captain goes down with the ship' (Taleb, 2018). A more detailed discussion from the same author: 'What is Skin in the Game? The central attribute is symmetry: the balancing of incentives and disincentives, people should also be penalized if something for which they are responsible goes wrong and hurts others: he or she who wants a share of the benefits needs to also share some of the risks'.

We will hold to the idea that symmetry of risk to reward is critical for skin in the game. If there is no risk, then there is no skin in the game. If the risk is too high, then there is no game as the agents will not be inclined to participate. The ratio of potential loss (L) to gain (G) will thus be a part of our metric.

Skin in the game also involves time. Temporal asymmetries are not explicitly considered for some endeavours because they are negligible. An example is day trading. If a day-trader hypes buying into a commodity at the start of trading, then by the end of the day verification will come along with associated gains or losses. A temporal asymmetry exists but is too small to shift skin in the game (i.e. $\tau_v \sim \tau_h$). Hyping a science result, on the other hand, allows for larger temporal asymmetries. This relates to the fact that not all science claims are created equal when it comes to verification/validation: In the words of Karl Popper (1962), '… some theories are more testable, more exposed to refutation than others; they take, as it were, greater risks'.

A motivating factor for hype, for advertising a result as significant, is some form of gains. If a hyped result is indeed a breakthrough, that will bring the largest of rewards. For a claim/result to be scientific, that requires verification/validation and the time to that is τ_v. Significant rewards can, however, be accumulated on a shorter time scale. Science has rewards in the form of career advancements, awards and funded grants. Perceived impact of a scientist's research feeds into those rewards and is a reward in and of itself. A measure of impact comes from how often a research paper is cited. Citations counts update monthly. Vitae/resume sections titled 'media contact' indicate that exposure in the public sphere also provides an impact measure (weekly updates from university press departments are telling in this regard). For science writers, the primary reward is readership. Maintaining or increasing readership operates on the time scale over which a journalist produces articles, which is generally weeks to months. Loss from a hyped result comes if it is shown to be invalid, the time to which is τ_v. That time scale affects potential losses in another way related to community memory. The longer τ_v, the more likely that original hype will dissipate and any study that invalidates the hyped result will tend to get less exposure.

The essential aspect of the above is that gains can come shortly after hype while potential loss is delayed. Figure 1 shows how this time asymmetry feeds into skin in the game. The graphs of Fig. 1 illustrate qualitative differences between hype scenarios. Different scenarios allow for different accumulations of benefits from hype (gain) and different potential down-sides for claims that do not live up to their hype (loss). Situations with a large temporal asymmetry, Fig. 1(a), should not be considered as equivalent to situations with temporal symmetry, Fig. 1(b). In the phrasing of systems design: There is benefit in
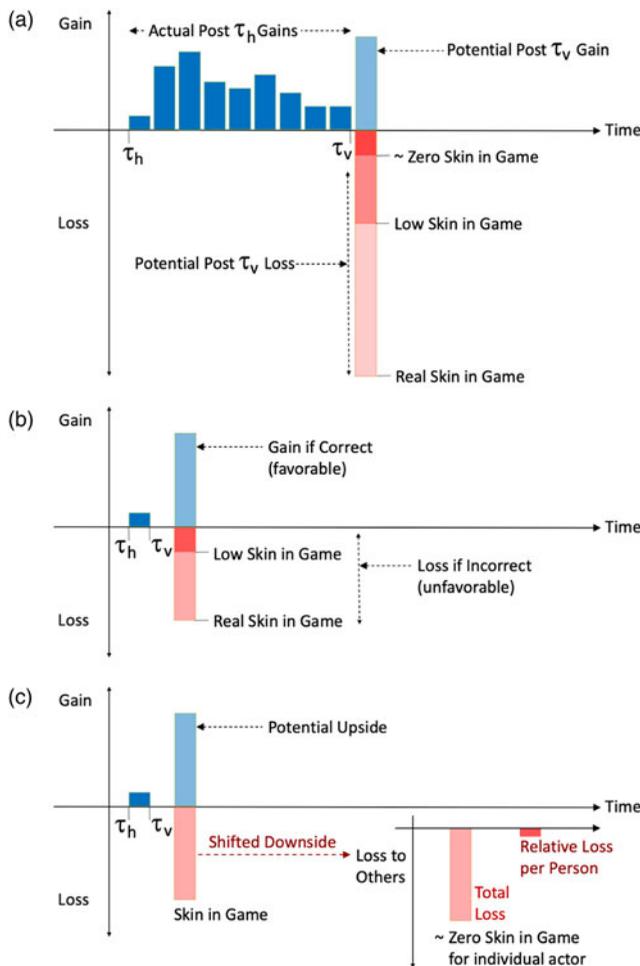
Fig. 1. *Illustrative graphs of potential risk to rewards evolution for hyped results with a long a verification time (top), a short verification time (middle), and distributed risks (bottom). The graphs highlight qualitative differences between hype scenarios. The gain bars represent benefits from hype that are accrued over finite time increments. This can include quantitative metrics (e.g. added citations per month) and more qualitative gains (enhanced prestige within a scientific community and/or at a university). The potential loss bars represent negative impacts if a claim does not live up to its hype. This includes community notoriety for over-selling research results to colleagues and/or the public. It also includes a deterioration of public trust in science (which is a distributed loss as it effects all scientists). An added asymmetry can result for claims/results that fall under the category of forecasts. Hyping a claim/result is, in effect, a forecast that it will be significant. If what is hyped is itself a forecast, then this leads to a doubling down without necessarily doubling the risk of losses.*

accelerating the benefit stream even if total rewards remain constant (Hazelrigg, 1996). Together with average loss to gain (risk to reward), this allows us to define a skin in the game metric (S) as

$$S = \frac{L\,(\tau_h - \tau_i)}{G\,(\tau_v - \tau_i)}. \tag{2}$$

where τ_i is the initiation time of a study. Together with our hype metric, this indicates that results/claims that come with a long τ_v can have low skin in the game and high hype factors. Conversely, more skin in the game, τ_v closer to τ_h, tends to lower the hype factor.

Some fields tend to have high skin in the game and low hype potential. An example is engineering fields where something is designed and built to be functional (one can hype a design before releasing it but the proof in the pudding will come soon after). Astronomy, astrobiology and planetary science tend toward the other end of the spectrum, as many claims/results only allow for validation by mission observations that are decades in the future. Results can be verified if other groups reproduce or refute them but that time scale is also long relative to the time scale of gains/rewards via citations and exposure. If the loss associated with a hyped claim being invalidated is low, then there is little individual skin in the game. This can occur if losses are distributed (Fig. 1(c)). It can also occur via an asymmetry in the advertising of positive relative to negative results (e.g. a study that invalidates a hyped claim). This connects pushed hype to pulled hype as science writers/journalists are less inclined to promote negative results as the rewards of readership tend to be lower.

The *S*–*H* spectrum is continuous but where a field or claim sits relative to end-members can be determined via a heuristic. Science careers last 10's of years. Rewards from hype come on an annual scale or shorter. If verification/validation operates on a rewards scale, then we are nearer the High-*S* and Low-*H* end of the spectrum. If it operates closer to a career scale, then we are nearer the Low-*S* and High-*H* end. Long time scale predictions/forecasts may not allow for validation on a career scale. If such predictions do not generate a desire from other groups to replicate, assess, or refute them, then they move toward zero skin in the game (this allows for claims that are hyped as significant to the public but generate no interest from the research community to confirm or refute them). The heuristic also applies to 'retrodiction' – studies that seek to explain historical/past events and/or observations. Just as the future is unwritten, the past is missing pages, which enhances uncertainties (some hyped results relate to conditions on planets more than a billion years ago). Greater uncertainties increase verification/validation times. As with predictions, a result also needs to spark other groups to confirm or refute it and the tools/data to do so must be available on a rewards time scale.

We re-stress that our discussion is specific to hyped science. Non-hyped research has skin in the game but of a different type. Researchers invest time and energy into projects and they look to re-coup that investment by presenting results to peers for evaluation, discussion and debate. The time and energy invested is compensated as the researchers are paid and funding for equipment/supplies comes via grants and/or institutional support (the exception would be a self-funded scientist – a case we do not address). This can proceed hype-free until a result is verified. The result could then be broadly promoted with a hype factor being zero and skin in the game staying near unity (high symmetry of risk to reward).

To summarize, skin in the game (*S*) is low for fields, or individual claims, that have a long time to verification/validation (τ_v) compared to the time that rewards are accumulated subsequent to hype. This allows for an inverse relationship between skin in the game and hype. Low skin in the game, and high hype potential, for any individual scientist, group, and/or journalist is not, in and of itself, detrimental to the scientific endeavour. To consider the effects on science as a whole requires thinking about the structure of collective science.

Structure and stability of collective science

The structure of science, from the time scientific societies formed and results became shared via publications, has been one of cooperation with competition. The cooperative framework encompasses sharing ideas, debating, critiquing, and acknowledging the work of others, all with the goal of coming to a collective understanding. The ideal of a pure collective, free of self-interests, is just that, and mixed in with cooperation there has been, and always will be, competition. Humans (e.g. scientists) can be motivated by wanting humanity to come to the right answers and be motivated by wanting to be the one who comes up with a right answer. Collective ideals interact with self-interests.

Scientists range from being more or less competitive and can cycle between cooperation and self-interest. Science is not, at present, at the self-interest end-member where every result is hyped. Cooperative incentives, internal and/or external, have kept it nearer the cooperative ideal. That said,

an element of competition has proved healthy in generating new ideas. That collective science does not fall into a singular mode is not unusual. Systems with interacting components often allow for multiple modes of behaviour with dynamic variations between them. We can illustrate this with a simple game theory model.

Our choice of a reference model state is one that allows for the co-existence of end-member strategies, hype or not hype results, and a stable mixed strategy where scientists only hype a certain proportion of results. We will choose the simplest model that allows for those conditions. Starting from the reference state, the model is designed to illustrate how changing incentives for hype can alter the structure of collective science. The values we will apply for rewards that come with choices to hype or not hype are examples only, and we make no claim they represent actual reward levels. The exact values will not be critical to our main point that changing incentives can lead to a structural change.

The model we envision is not a one-time game. It's played every time a player is faced with a hype or non-hype decision. For simplicity we will frame the model as a two-player game. However, how any single player views the decisions of others allows for different views of what a player represents. For example, if player two is a single scientist making a decision, then he or she will not view player one as some other single scientist. Rather, they will consider the general tendency of their collective community at that time. Player one is then viewed as, in effect, the background environment of hype versus non hype for a community (e.g. are the overall tendencies at the time leaning more toward hype or non-hype).

Figure 2 shows a set of payoff matrices for a game theory model of cooperative science mixed with competitive self-interest. The farthest left column, for any of the matrices, represents the choices/actions of player one while the top row represents those of player two. The two number sets within a matrix represent payoffs under different scenarios with the first number being player one payoff and the second player two payoff. The matrix in the top left corner is the reference state model. The others represent payoff changes due to shifting incentives and/or disincentives.

The reference model of Fig. 2 allows for three Nash equilibria – scenarios under which each player remains indifferent to changes in the others strategy (Spaniel, 2021). Scenarios where both players hype or do not hype are pure strategy equilibria. If both players fall in either, then they have no incentive, no added payoff, to change strategy if the other player changes his/her strategy. There is also a mixed strategy equilibria in which each player hypes a certain proportion of results. To see this, we can consider the proportion of time a player hypes a result. If that is given by σ, then $(1 - \sigma)$ is the proportion of non-hyped results.

Consider player one playing a mixed strategy. A required condition for equilibrium is that player two must be indifferent to playing either of their pure strategy options in the face of player one's mixed strategy. Indifference means that player two's expected utility, *EU*, will be the same for either strategy. Consider player two choosing to not hype, *N*, while player one plays a mixed strategy. Player two's
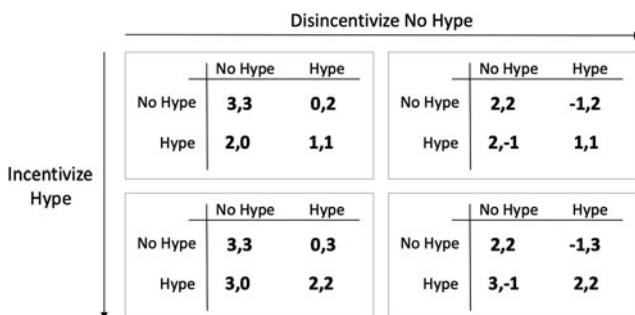

Fig. 2. *Payoff matrices for collective science under shifting incentives and/or disincentives regarding hype.*

expected utility, EU_N, will be

$$EU_N = (1 - \sigma)(3) + (\sigma)(0). \tag{3}$$

Now consider player two choosing to hype, H, for the same mixed strategy of player one. Player two's expected utility, EU_H, will be

$$EU_H = (1 - \sigma)(2) + (\sigma)(1). \tag{4}$$

To leave player two indifferent, player one's mixed strategy must be such that

$$EU_N = EU_H. \tag{5}$$

Solving for σ leads to the conclusion that in order for player two to remain indifferent, player one's mixed strategy allows them to hype 1/2 of their results. The symmetry of the payoff matrix (Fig. 2) leads to the same conclusion for player two choosing to play a mixed strategy. Mixed strategies in a population can reflect the distribution of players playing pure strategies. Thus, a single scientist viewing the background tendency of a community as, in effect, another player, will consider the proportion of the community tending to hype at a given time.

The above can be generalized for variable payoffs. The payoffs for player two not hyping while player one does not or does hype are denoted by P_{NN} and P_{NH}, respectively. The payoffs for player two hyping while player one does not or does hype are denoted by P_{HN} and P_{HH}, respectively. A mixed strategy Nash equilibrium exists provided that

$$(1 - \sigma)(P_{NN} - P_{HN}) = (\sigma)(P_{HH} - P_{NH}). \tag{6}$$

Incentives to hype and/or disincentives to not hype lead to progressively lower values of σ that allow for a mixed strategy equilibrium. The top right and bottom left matrices show that σ must decrease if all players are to remain cooperative in the sense of feeling no pressures to switch strategies. If pushed far enough, incentives to hype and/or disincentives to not hype can lead to the loss of a mixed strategy equilibrium. The right bottom payoff matrix of Fig. 2 shows how a mixed strategy can become unstable as σ is zero under those payoffs, i.e. what was once a mutually favourable strategy is no longer an equilibrium. A change in the number of equilibrium states within a system is a hallmark of structural instability (Guckenheimer and Holmes, 1983).

Although σ may need to decrease to maintain equilibrium, under shifting conditions, that does not mean that it will. Consider player two as an individual who notices that the community tendency for hyping results remains fixed even though payoffs have shifted (Fig. 2). The individual will no longer be indifferent. Expected utility would be higher if they chose to preferentially hype. This occurs even if the maximum reward for all players would occur under a collective no one hypes scenario. Player two is not motivated by selfishness nor is the community as whole. That is to say, there does not need to be any bad actors in the system to drive it toward instability. Some players simply hold to a strategy that has worked. Others see a group tendency starting to lower their relative rewards unless they adjust (the motivation is not selfish lust for more gains but simply not wanting to fall behind and risk being removed from the game all together).

Pressure on an individual, due to increased hype, can be considered via a simple model of cooperation decline (James, 2012). If every result is hyped, then science moves toward a mode where everyone is competing for 'air time'. What prevents that, and maintains cooperation, is a level of cost for overhyping. That is a function of skin in the game, and we can express it as $C(S)$. Working against this is the proportion of scientists, P, that are hyping results. Provided that cost outweighs that by some factor k, then a non-hyping individual will not be pushed to change. The condition for

maintaining cooperation can be written as

$$C(S) - kP > 0. \tag{7}$$

Different scientists will have different tolerances, i.e. different values of k. However, they will also have a limit at which they start to feel that by holding to a non-hype strategy they are, in effect, falling behind, particularly if those who are hyping are gaining rewards with little to no risk. This can cause a shift of strategy. As more non hype strategists switch over, the rewards of non-hype deteriorate. This drives the system toward structural instability (Fig. 2). Stated another way, small instabilities within the system (individual non hype strategies becoming unstable) can move the system itself toward a structural instability (death by multiple small cuts). The conditions that lead to that depend on skin in the game, S. Low S lowers the cost of hype and, as a result, increases the potential of structural instability. In effect, low S accelerates a positive-feedback. Negative feedbacks, that act to maintain and/or restore an equilibrium, can be operative but if they operate slower than the positive feedback, then the system will still move toward a structural change (Dorner, 1996). Self-correcting feedbacks, that maintain cooperative science, operate at the community level. That requires levels of coordination (and group motivation) that can make them move relatively slow.

Although instability can occur with no bad actors, the influence of overly competitive individuals can have stronger effects if skin in the game is low. Every scientist has a story about an individual who reviewed their paper(s) and insisted that his or her (usually his) work be extensively cited. We have encountered a referee who insists that citations to work that contradicts his own not be cited. There are scientists who actively seek media exposure. Every field has at least one established scientist who 'argues from authority' on media platforms. Individuals prone to self-promotion have always existed and always will. They are a minority. However, a vocal minority can have more effect than raw numbers might suggest (vocal is a good description for those who will always be drawn to having their views voiced louder, i.e. hyped). Fields with high skin in the game are resistant to that influence as $C(S)$ can outweigh it. On the flip side, low skin in the game allows a vocal minority to have a greater influence.

Acknowledging the potential of bad actors, we re-stress that instability can occur even if the majority of scientists do not act out of selfish interests. A drift toward instability can also be initiated by factors external to working scientists. We take this up in the next section.

Perturbations and nudges toward instability

Like other systems, collective science has evolved to develop structure and hierarchy (Bejan, 2019). External perturbations can initiate system restructuring. External factors can be well intentioned yet have detrimental effects (Dorner, 1996). No factor needs to be large in immediate effect to initiate a drift toward structural change. Fields with lower skin in the game can reflect a structural drift sooner than others, i.e. lower skin in the game leads to less damping of perturbations.

An example of a perturbation is the rise of university public relations departments dedicated to advertising scientific results via press releases, social media and/or interviews. This can promote science but the main goal is to promote research from the university itself. This creates a pull toward hype. Scientists feel the pressure to produce 'media worthy' results that will be picked up by their public relations department. The associated rise of vitae sections titled 'media exposure' is telling, as is the level to which scientists now announce research results on social media platforms. The nudge can start early as many PhD theses now also contain sections documenting the candidate's popular science media coverage.

Improving public understanding of science is a worthy goal. It has led to training sessions at international meetings. Helping scientists communicate clearly and effectively is one thing but it often gets shifted toward making ideas seem more interesting and/or providing a level of entertainment. The assumption is that everyone's research can and should be made interesting to the public, i.e. science should not be 'boring'. Some science, if presented clearly, will be boring to the general public. Framing that as a problem to be corrected is a nudge toward hype. The idea that scientists should
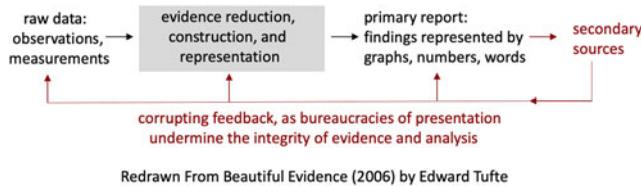
Redrawn From Beautiful Evidence (2006) by Edward Tufte

Fig. 3. *Redrawn from the section titled 'When Evidence is Mediated and Marketed: Pitching Out Corrupts Within' (pages 154–155 of the cited work). Since Edward Tufte's critique, primary reports makers have been more incentivized to push their results toward secondary reports (i.e. it has come to be perceived as a source of career rewards).*

develop skills for upping the public interest factor increases the nudge. It also introduces a new level of competition: 'my idea is of value within my field but will it be interesting beyond it?'. This connects to the idea of broader impact.

Improving the societal impact of science is a worthy goal for science as a collective. It is a different matter if that is turned into criteria for individual research funding. Moving funding decisions beyond a competition between scientific ideas, toward something broader, is a nudge. One interpretation of 'broader' is impact beyond scientific field. Many universities now aid scientists with broader impacts. Part of that relates to how results will be disseminated beyond primary source platforms. The downsides of that feedback have been well articulated (Tufte, 2006) (Fig. 3).

Associated with the above is the rise of funding criteria for 'transformative' research. Pressure to justify research as transformational is a nudge toward hype that downplays confirmation studies (confirmation and transformation are viewed as different things). As a result, the time to verification/validation gets pushed toward larger values, skin in the game goes down, hype factors increase, and the system becomes more exposed to instability.

Science has always had competition for jobs. More recent are workshops on how to get a job that include 'pitching your science'. This has been augmented by a rise of early career awards and their value for career advancement – a nudge away from cooperative science. Established scientists also succumb to that nudge and often initiate award nominations for colleagues at their own institution. The detrimental effects of awards, on collective science, were noted some time ago (Merton, 1968). The number of awards has increased since then.

Many journals now send authors links on how to increase the visibility of accepted articles or, to use wording we have received, to get ones' research 'the visibility it deserves'. The tips include methods on promoting and branding ones' research. The message sent is that self-promotion of published papers is required to remain competitive (a variant of the red queen dilemma in that one needs to take added measures not to advance but to not fall behind).

The above is not inclusive. We have left off more obvious nudges (e.g. predatory journals, popular science writers/bloggers who need to fill daily content, citation metrics and altmetrics to evaluate impact together with the gaming of metrics). Our intent is to show that there is no shortage of nudges. The nudges and reactions of individual scientists can all be free of ill intent and still move the cooperative system of science toward instability. The structure of the system itself allows for a 'tragedy of the commons' scenario.

Conclusion, discussion and potential actions

Our motivation was the question of 'Why Do So Many Astronomy Discoveries Fail to Live Up to the Hype?' (Falk, 2021). We expanded it to address the rise of hype in science. We argued that the structure of collective science exposes it to a hype instability. Low skin in the game allows nudges to initiate internal feedbacks and a drift toward structural instability, i.e. a system reconfiguration. Along the drift, the number of results hyped as breakthroughs will increase and more claims will fail to live up to the hype. This provides a proposed answer to the motivating question.

Our motivation was not a search for fixes to a hype problem, beyond the idea that if one is going to address a systemic issue one needs to start with the structure of the system itself. That said, we can offer some thoughts on moving forward.

If a system is moving toward a structural shift, then there are three courses of action: Intervention, Reset, No Action. Intervention introduces new nudges/constraints designed to move the system away from a tendency perceived as harmful. Reset seeks to remove or damp nudges/constraints and internal feedbacks that set off a drift toward structural change. No Action is motivated by the idea that a structural shift is inevitable and/or that it may be for the good (structural instability is not a bad thing if the new structure is better than the old).

It could be argued that there is no problem to be dealt with (Falk, 2021). The rationale is that different groups promoting their ideas are healthy for the generation of new ideas. We agree for hype-free science, where hypotheses have been debated for as long as collective science has existed. The issue is: Does it apply under hype conditions? Some scientists may agree that enhanced competition generates more novel ideas. Others may hold that even if hype is not beneficial, it does little harm. We address each in turn.

Competition, co-existing with cooperation, has proved healthy for the generation of scientific ideas. Hype can move science toward the more competitive end. Will competition remain healthy under that shift? Increased competition cannot lead to a continual increase in quality ideas. Consider the limit where competitive interests dominate over cooperative ones. The number of ideas might increase but few would argue that quality will (Edwards and Roy, 2017). The question then becomes is the system close to a roll-over point? That is debatable. Our point here is that the perceived value of enhanced competition is not a strong argument for no action unless proponents can argue that the system is far from a roll-over.

Even if competition can catalyse new ideas, hype generated debate will play out in a different forum. Seeing more scientists announcing more results in the public sphere, in order to get their ideas noticed, will affect the public view of science. The prediction that the drive to get things noticed could redefine value has played out to be prophetic (Goldhaber, 1997). More and more scientific claims vying for attention (being hyped) can shift the public perception of what scientists, and science itself, values. Science becomes yet one more entity vying for attention. Unlike the hype-free forum, the attention vied for does not represent the balance of science. Confirmatory studies are not viewed as worth hyping (old news) and contradictory studies rarely get broadcast. A warning about the latter exceeds the warning about attention redefining value by almost 400 years: 'It is a peculiar and perceptual error of the human understanding to be more moved and excited by affirmatives than by negatives' (Bacon, 1620).

We will add one more counter to the idea that hype does no harm. Hype works against diversity. Many early-career scientists will not be comfortable with enhanced competition and signals that the value of ideas will fall short if they do not develop skills to broadcast those ideas. Increased rewards for the competitive/hype side of science will affect the number of people coming into the system who are drawn to the cooperative side. That does not open the door for diversity. Hype can shut that door in another way. It is universities with resources that have the avenues to hype results. This works against less established universities with lower resources. Those universities draw a diverse student pool.

Our view is that no action is not the best action. That said, caution should be taken with reactions designed to 'correct' an unfavourable situation. Addressing systemic issues starts with the structure of the system itself. Understanding the structure can highlight problems with what appear to be reasonable interventions.

Circling back to the article that motivated this one, a topic called out within it was the search for life beyond Earth (Falk, 2021). The astronomy/astrobiology community has felt some embarrassment from claims of extraterrestrial life that do not live up to hype. This has led to ongoing discussions of standards and regulations for how research into that topic should be disseminated (Green *et al.*, 2021). The motivation is well intentioned but fails to appreciate the structure of the system.

Applying a 'correction' to a system with multiple equilibria, in an effort to undo previous effects, will not get one back to the same starting point. From a game theory perspective, the history of how a game has

been played can have a dominant influence on the way it will play out. Interventions cannot erase history and can often have effects as detrimental as the original nudges they are meant to negate (especially when those nudges remain in place). More generally, this approach fails to appreciate the systemic problem from the start. Hype, low skin in the game, and enhanced competition goes beyond a single topic. Focusing on a single topic (often the most obvious one) will not stop a drift toward instability. Similarly, trying to fix the consequences of some actions will not get at their root causes. Applying corrections and/or interventions also misses the true value of skin in the game. Skin in the game is not about restrictions. It's about balance and fairness when one's actions have the potential to affect others (Taleb, 2018).

The above also highlights problems of focusing on the state of a system at a particular time. This leads to strategies guided by regulating situations as opposed to processes, with associated temporal evolutions and time lags. The end result is regulations that 'oversteer' the system, creating new problems down the road as they seek to correct current situations (Dorner, 1996). The rise of hyped claims for life beyond Earth did not result from a lack of regulations about how research into extra-terrestrial life should be disseminated. Imposing new regulations/standards will not get at the cause of the rise. It could also be viewed as suppressing alternate ideas. The downsides of that would be significant and would appear with a shorter time lag than any upsides new regulations might have.

As well as tending to miss root causes of structural shifts, interventions add constraints. Constraining systems can restrict their freedom to develop functional hierarchies (Bejan, 2019). This does not mean that intervention is not an option but, in our opinion, it should not be the first consideration. That leaves reset. It would be nice if reset meant remove nudges and watch the system drift back to a favourable mode of behaviour. Systems with multiple equilibria and path dependencies are not part of that nice and simple world. Reset needs to be viewed in context.

It is easy to attribute system drift to external perturbations and/or the action of bad actors. That is, to place the blame on 'the environment' or 'the other'. The first step, if reset is to be effective, is for the community to get past those tendencies. At the risk of over personalizing, we start with a self-critique.

When we first discussed the question of why hype has increased, we fumbled for an 'answer'. We talked past each other by assuming we all had the same meaning in mind for terms like 'hype' and 'skin in the game'. We threw those terms around to provide simple answers to a simply posed question. We blamed overly self-motivated scientists and suggested regulations to reign them in. As we tried to be more precise, we saw the flaws in our thinking. We realized how easy it is for participants within a system to miss the structure of the system and to mis-judge the effects of their actions within it. We realized how easy it is for working scientists to respond to even the slightest of nudges toward hype and then provide internal justifications ('it's part of moving my career forward'; 'if I don't broadcast my ideas, while others do, I will fall behind'; 'it's just a new job requirement'; 'what's the harm'). In short, we realized we are part of the problem. When we circled back to the article that motivated us, we saw sign-posts we had missed. Prominent scientists and science writers fumbled for answers, gave single phrase answers, or noted that the issue is complicated. They used the terms 'hype' and 'skin in the game' with little effort at precision. None of the interviewees, or the writer of the article, acknowledged the possibility that they are part of the problem. We re-stress, we made the same mistakes (more likely to larger degrees).

Our point above is that although hype is discussed, it does not garner deep thought from the community at large. Science organizations have time for career training sessions on pitching science and interacting with the media but little time for discussions about the direction and/or stability of cooperative science. It may well be assumed that education into those aspects of a scientist's career come in the classroom and/or that, for established scientists, they are no longer a concern. As a result, efforts at coordination become unbalanced. The cooperative/collaborative side of science is viewed as too self-evident to require coordinated efforts to set expectations. That, in our opinion, is what needs to be reset before any other resets can have any effect. For systems with multiple equilibria, coordination of expectations can influence which equilibria a system will move toward (Krugman, 1991). Coordination requires an internal reset amongst participants within the system versus a focus on regulations, nudges and/or environmental circumstances.

To gauge whether reset is possible, we can pose a question to research scientists: How difficult is it to say 'no thanks' when a reporter contacts you to do a write up about your new results? It should be easy yet, speaking for ourselves, it is not. A step toward reset is to ask why is it not easy. Is the only issue the reporter, or a public relations officer, coming to you? Do you, as a working scientist, feel you have no real control regarding the answer if you want your career to thrive? How much have you bene-fited from hype? How much has your field of science benefitted from collective efforts to hype its value? How much have you been involved in hype, directly or indirectly, and at the same time argued that science needs to be more inclusive and enhance diversity? Do you see any downsides to broad-casting your results before they have been verified/confirmed? Do you feel any of this is worth discus-sion amongst your colleagues? If not, then a reset is unlikely. Will that cause a collapse of science? No. Science can exist at the competitive end of the spectrum where all scientists vie for attention (structural instability does not delete the system itself). It will be just a different form of science.

Ending this article at this stage relates to its motivation. The motivation was not to propose solu-tions. It was to answer a hype related question treating it, as best we could, free of judgements as to how important or trivial it may be, up until this last section. Our suggestion, if the community agrees that hype is a problem, is to give discussion about it some space. Contact journal editors and meeting organizers. Propose sessions. Talk to administrators who create pressures for hyping your results. Talk to programme supervisors. Take time to research and think about what skin in the game, rewards from hype, and strategic interdependence mean for your field and for you as a working scientist. Realize that the collective system of science is principally the collective of working scientists and not externals that may create nudges in certain directions. That brings the potential for collective power and individual responsibility for the health of the collective system.

References

Bacon F (1620) Nuvum Organum (English translation). In Ellis RL and Spedding J (eds). *The Philosophical Works of Francis Bacon*. London: Routledge, pp. 212–387.

Bejan A (2019) *Freedom and Evolution: Hierarchy in Nature, Society, and Science*. New York: Springer.

Dorner D (1996) *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*. New York: Metropolitan Books.

Edwards MA and Roy S (2017) Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and hyper-competition. *Environmental Engineering Science* **34**, 51–61. doi:10.1089/ess.2016.0223.

Elmore SA (2018) The altmetric attention score: what does it mean and why should I care? *Toxicologic Pathology* **46**, 252–255.

Falk D (2021) Why do so many astronomy discoveries fail to live up to the hype?, Undark Magazine. Available at https://undark.org/2021/01/18/astronomy-discoveries-fall-victimto-hype.

Funtowicz S and Ravetz J (1990) *Uncertainty and Quality in Science for Policy*. The Netherlands: Kluwer Academic Publishers.

Goldhaber MH (1997) Attention Shoppers: The Currency of the New Economy Will Not Be Money, but Attention: A Radical Theory of Value, Wired Magazine.

Green J, Hoehler T, Neveu M, Doumagel-Goldman S, Scalice D and Voytek M (2021) Call for a framework for reporting evidence for life beyond Earth. *Nature* **598**, 575–579.

Guckenheimer J and Holmes P (1983) *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer.

Hazelrigg GA (1996) *Systems Engineering: An Approach to Information-Based Design*. NJ: Prentice-Hall.

James A (2012) *Assholes: A Theory*. New York: Doubleday.

Krugman P (1991) History versus expectations. *The Quarterly Journal of Economics* **106**, 651–667.

Merton RK (1968) The Matthew effect in science. *Science (New York, N.Y.)* **159**, 56–63.

Popper KR (1962) *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Basic Books.

Spaniel W (2021) *Game Theory 101: The Complete Textbook, Copyright William Spaniel*. CreateSpace Independent Publishing Platform.

Taleb N (2018) *Skin in the Game*. New York: Random House.

Tufte ER (2006) *Beautiful Evidence*. Chesire, CT: Graphic Press.

Warren HR, Raison N and Dasgupta P (2017) The rise of altmetrics. *JAMA* **317**, 131–132.