NBER WORKING PAPER SERIES

EQUILIBRIUM CONDITIONS FOR CATCH-22 SITUATIONS

Joshua S. Gans

Working Paper 33304
http://www.nber.org/papers/w33304

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2024

Joshua Gans has drawn on the findings of his research for both compensated speaking engagements and consulting engagements. He is also chief economist of the Creative Destruction Lab, a University of Toronto-based program that helps seed-stage companies, from which he receives compensation. He conducts consulting on cryptocurrency and payments matters with an association with Keystone and his ownership of Core Economic Research Ltd. He also has equity and advisory relationships with a number of startup firms. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33304

Equilibrium Conditions for Catch-22 Situations
Joshua S. Gans
NBER Working Paper No. 33304
December 2024
JEL No. C72, D82

ABSTRACT

This is a paper in the ``economists ruin everything" field. It considers whether Catch-22 situations can persist as an equilibrium phenomenon. Rather than being an arbitrary rule or a set of self-serving beliefs, the focus is on the preferences of Gatekeepers who choose to create such situations in the first place. The base game-theoretic model is of a Catch-22 situation inspired by Heller's famous paradox. We consider a Requester who may be Sane or Insane and a Gatekeeper who must decide whether to grant the Requester's desired outcome or force them into a less desirable one. This is modelled as a game in which the Requester chooses whether to send a request signal before the Gatekeeper decides. We solve for the conditions under which a Catch-22 situation persists as an equilibrium and its efficiency properties. It is demonstrated that Catch-22 situations can arise, but they reflect an efficient response on the part of a Gatekeeper facing asymmetric information. An application to labour markets is also considered

Joshua S. Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
and NBER
joshua.gans@rotman.utoronto.ca

> *“You mean there’s a catch?”*
>
> *“Sure there’s a catch,” Doc Daneeka replied. ”Catch-22. Anyone who wants to get out of combat duty isn’t really crazy.”*
>
> *There was only one catch and that was Catch-22, which specified that a concern for one’s own safety in the face of dangers that were real and immediate was the process of a rational mind. Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions. Orr would be crazy to fly more missions and sane if he didn’t, but if he was sane, he had to fly them. If he flew them, he was crazy and didn’t have to; but if he didn’t want to, he was sane and had to. Yossarian was moved very deeply by the absolute simplicity of this clause of Catch-22 and let out a respectful whistle.* (Heller, 1961)

1 Introduction

Joseph Heller’s classic *Catch-22* portrays a paradox: a pilot can only be relieved from flying dangerous missions if they are insane. But any pilot who requests relief, by that very action, reveals a concern for self-preservation, implying sanity. Thus, to qualify for relief, one must never request it, but not requesting fails to demonstrate insanity.

This is a canonical Catch-22 situation: an agent cannot access a rule because complying with the rule prevents them from achieving their desired outcome. In popular culture, it presents a type of ‘no win’ situation or dilemma but of a particular kind. Specifically, the rule is set by those in power that ultimately leaves others with no real choice. It connotes a situation where such rules result in oppression, and their rationale subverts outcomes other than those pushed by those in power. Consequently, Catch-22 situations have been argued to be present in real-world situations, e.g., “to receive a loan, you need a credit score, but to obtain a credit score, you need a loan.”

Economists, as they are wont, nay have a duty, to do, might wonder: where did these rules come from in the first place? Are they a mere expression of power dressed up in a rationale or did they express the preferences of those who imposed them? This applies in Heller’s scenario. There, the authority (Gatekeeper) expresses their own preferences: they do not want to ground rational (Sane) pilots who should remain in action, but they do want insane pilots to be kept from flying. This structure suggests we must model not only the paradoxical condition but also the payoffs to both sides. Only then can we understand why the rule is in place, what its function is and whether it achieves that function. Only then do we have a theory to be put against the alternative notion that the rule exists because of guile, comedic intent or existential angst.

We formalize the game as follows. There are two players: a Requester, who can be one of two types (Sane or Insane) and a Gatekeeper. The Requester chooses to send a message (Request or Not Request) before the Gatekeeper decides whether to grant their request or not. The Requester's sanity is defined by their preferences: Sane types prefer A (not to fly) to B (to fly), while Insane types prefer B to A; indeed, their preferences define their insanity. The Gatekeeper's preferences run counter: the Gatekeeper wants Sane types to get B and Insane types to get A. In Heller's terms, the reason the Gatekeeper does not want insane agents to fly is that they would prefer them not to fly.

This game allows us to explore the two characteristics that underpin a Catch-22 situation. First, the agent subject to the rule *observe*s its futility in practice. In Heller's case, their request not to fly will always be denied. Second, the Gatekeeper draws clear *inferences* from the Requester of their type based on the message they send (i.e., the request they do or do not make). In analysing this game, we look for equilibrium outcomes that are observationally a Catch-22 situation in that the rule is applied and also ones that are underpinned by the beliefs associated with a Catch-22 situation.

Section 2 sets the game up, assuming that the Requester moves first and the Gatekeeper responds. Section 3 then solves for the equilibrium outcome. It is demonstrated that a Catch-22 situation – that is, where a Sane type who requests not to fly will have to fly – only arises if the portion of Sane types in the population is relatively high with a threshold related to the Gatekeeper's costs of allowing an insane person to fly or preventing a sane person from flying. When there are many Insane types, the Gatekeeper is better off allowing sane types to request and fulfil their wish not to fly and barring anyone else from flying on account of them being insane. The reverse timing, where the Gatekeeper commits to a rule ex ante, is shown to give rise to the same equilibrium outcomes.

In Section 4, we then turn to consider a more economically relevant Catch-22 (conjectured) situation: "You can't get a job without experience, but you can't get experience without a job." This is done in the context of a labour market where getting a job gives you experience and the ability to perform (and earn) based on higher productivity. It is shown that ...

A final section concludes with thoughts about this general class of Catch-22 situations with a view to discouraging any others from wasting more time pursuing this line of research.

There is a small literature on Catch-22 as it relates to economics. It pops up to motivate a few policy papers, including Colgan and Quinlin (1997), de Rouffignac (2000), Flood and Toner (1997), Brams (1999) and Stein (1999). Brams and Jones (1999) explores some game theoretic explorations in the context of 2x2 games and power while Carmichael (2002) considers a potential Catch-22 situation in a signalling game where sending the desired message to one receiver results in the wrong message to the other and vice versa. None of these papers, however, have tackled the Catch-22 situation as spelled out by Heller (1961) head-on nor explored its plausibility for that case.[1]

2 Model Setup

There are two players: a Requester and a Gatekeeper. The Requester is either Sane (S) or Insane (I). Let $p \in (0,1)$ be the probability that the Requester is Insane, and $1-p$ the probability that the Requester is Sane. There is asymmetric information as the Requester knows their type while the Gatekeeper does not. The Gatekeeper only knows p.

The Requester chooses between two messages: R (Request) or $\neg R$ (Not Request). After observing the Requester's action, the Gatekeeper chooses between G (Grant) or D (Deny). If the Gatekeeper chooses G, the outcome is A; if D, the outcome is B.[2] Thus, the timing of the game is as follows:

1. Nature chooses the Requester's type: S with probability $1-p$ or I with probability p.
2. The Requester chooses R or $\neg R$.
3. The Gatekeeper observes the Requester's action and chooses G or D.

The preferences of each player are as follows:

- **Requester's Preferences**: Let $r_i(a)$ be the payoff a Requester of type $i \in \{S, I\}$ receives with an outcome $a \in \{A, B\}$. It is assumed that $r_S(A) > r_S(B)$ (a Sane Requester prefers A over B) and that $r_I(A) < r_I(B)$ (an Insane Requester prefers B over A.
- **Gatekeeper's Preferences**: Let $g_I(a)$ be the payoff the Gatekeeper if the Requester is of type $i \in \{S, I\}$ and receives an outcome $a \in \{A, B\}$. For a Sane Requester, $g_S(A) < g_S(B)$ (the Gatekeeper prefers B over A).

Thus, sanity is defined by which outcome the Requester prefers: a Sane Requester wants A, an Insane one wants B.

---

[1] There is some work in logical philosophy on Catch-22 situations that argues that the proposed rule and inference are vacuous and uninformative (Goldstein, 2004). Here, I consider a game that allows for non-arbitrary rules, and any rule formation is consistent with incentives and beliefs. Thus, it provides a different perspective than this literature.

[2] An alternative timing where the Gatekeeper first commits to a rule is considered below with no change in the equilibrium outcomes.

3 Equilibrium Outcomes

We examine perfect Bayesian equilibria of the game; that is, a profile of strategies and beliefs where each player's strategy is a best response given their beliefs at each stage of the game, and beliefs are consistent with Bayes' rule whenever possible. For each type of Requester, a strategy is a probability ($\beta_S \equiv P(R|S)$ for Sane, $\beta_I \equiv P(R|I)$ for Insane) of choosing R and, for the Gatekeeper, a strategy is probabilities, $(\alpha_R, \alpha_{\neg R})$ of choosing G after R or $\neg R$. In equilibrium, (i) given the Gatekeeper's strategy, each type of Requester maximises their expected utility by choosing R or $\neg R$; (ii) given the Requesters' strategies, the Gatekeeper chooses G or D to maximise expected utility after observing R or $\neg R$ and (iii) the Gatekeeper's beliefs, $P(I|R)$ and $P(I|\neg R)$ are updated using Bayes' rule:

$$P(I|R) = \frac{p\beta_I}{p\beta_I + (1-p)\beta_S}, \quad P(I|\neg R) = \frac{p(1-\beta_I)}{p(1-\beta_I) + (1-p)(1-\beta_S)}$$

In some cases, if certain actions are played with probability zero, off-path beliefs must be specified.

Working backwards, after seeing R, the Gatekeeper's expected payoff from choosing G and D are respectively:

$$E[G|R] = P(I|R)g_I(A) + (1 - P(I|R))g_S(A).$$

$$E[D|R] = P(I|R)g_I(B) + (1 - P(I|R))g_S(B).$$

The Gatekeeper grants after R if $E[G|R] > E[D|R]$. Similarly for $\neg R$.

Define $\Delta_I = g_I(A) - g_I(B) > 0$ and $\Delta_S = g_S(A) - g_S(B) < 0$. Then:

$$E[G|R] - E[D|R] = P(I|R)\Delta_I + (1 - P(I|R))\Delta_S.$$

It is straightforward to see that the Gatekeeper will choose G rather than D following a request (R) if:

$$P(I|R) > \frac{-\Delta_S}{\Delta_I - \Delta_S}.$$

and will choose G rather than D following no request ($\neg R$) if:

$$P(I|\neg R) > \frac{-\Delta_S}{\Delta_I - \Delta_S}.$$

Otherwise, the Gatekeeper will be indifferent.

Turning to the Requester's choice given an expectation that the Gatekeeper chooses

$(\alpha_R, \alpha_{\neg R})$, the Sane type chooses R if it yields a higher expected payoff, that is:

$$r_S(R) = \alpha_R r_S(A) + (1 - \alpha_R) r_S(B), \quad r_S(\neg R) = \alpha_{\neg R} r_S(A) + (1 - \alpha_{\neg R}) r_S(B).$$

Note that Sane is indifferent if $\alpha_R = \alpha_{\neg R}$. Similarly, for an Insane type:

$$r_I(R) = \alpha_R r_I(A) + (1 - \alpha_R) r_I(B), \quad r_I(\neg R) = \alpha_{\neg R} r_I(A) + (1 - \alpha_{\neg R}) r_I(B).$$

Indifference implies $\alpha_R = \alpha_{\neg R}$ as well. So if both types must be simultaneously indifferent, $\alpha_R = \alpha_{\neg R} = \alpha$. Otherwise, if $\alpha_R \neq \alpha_{\neg R}$, both Sane and Insane Requesters will find a pure strategy optimal. If $\alpha_R > (<) \alpha_{\neg R}$, Sane chooses R $(\neg R)$ and Insane chooses $\neg R$ (R).

Given this, we can now prove the following:

Proposition 1 *Consider the game where the Requester moves first. Let*

$$q \equiv \frac{-\Delta_S}{\Delta_I - \Delta_S} = \frac{g_S(B) - g_S(A)}{(g_I(A) - g_I(B)) + (g_S(B) - g_S(A))}.$$

1. *If $p < q$, the unique equilibrium outcome has the Gatekeeper always denying.*
2. *If $p > q$, the unique equilibrium outcome has the Gatekeeper always granting.*
3. *If $p = q$, there is a continuum of equilibria where $\beta_S = \beta_I$ and $\alpha_R = \alpha_{\neg R}$.*

All proofs are in the appendix. When the probability of insanity p is low (less than q), the Gatekeeper expects mostly Sane types and, thus, prefers forcing B by denying. When p is high ($p > q$), expecting more Insane types, the Gatekeeper grants to achieve outcome A. Only at the exact balance point $p = q$ is the Gatekeeper genuinely uncertain and indifferent, leading to a mixed strategy outcome that is, arguably, non-generic.[3]

This demonstrates that a Catch-22 situation whereby the Gatekeeper always, observationally, denies a request for A from a Requester only arises if p is sufficiently low (less then q). This is because, in that case, the Gatekeeper's error from having an Insane person with outcome B (Δ_I) is lower relative to the error in having a Sane person with outcome A (Δ_S). Note, however, while the Gatekeeper is acting *as if* it is denying because it is inferring that only a Sane person would make a request R, the Gatekeeper's actual equilibrium beliefs that the person making the request is Sane is arbitrary. The Gatekeeper could hold those beliefs

---

[3] At $p = q$, can standard refinements like the Intuitive Criterion eliminate any equilibria? Since all parties are indifferent, no off-equilibrium signal changes payoffs significantly. The Requester cannot send a costly signal to shift beliefs beneficially. The Gatekeeper's best response is linear in $P(I|a)$. Thus, refinement arguments do not rule out the continuum of equilibria. The Catch-22 nature remains robust.

or not. It does not matter for the outcome. This is the classic babbling equilibrium outcome from Crawford and Sobel (1982).

More critically, the equilibrium outcome is driven by efficiency considerations. Absent an ability to infer an agent's type from cheap talk messages, the Gatekeeper's action reflects a balancing of the types of errors they could make. Thus, the rule is not arbitrary in this model but reflects the preferences of the Gatekeeper. Specifically, Proposition 1 shows that if there is a large volume of Insane types, the Gatekeeper would not employ a Catch-22-like rule.

Now consider what happens if the Gatekeeper commits to $(\alpha_R, \alpha_{\neg R})$ before the Requester chooses R or $\neg R$. The Requester, observing $(\alpha_R, \alpha_{\neg R})$, picks the message that maximises their payoff. If $p < q$, any attempt to equalise payoffs by mixing is exploited by Requesters. The Gatekeeper ends up best off by choosing a pure strategy. The same logic applies for $p > q$. Thus, we get the same outcomes: always deny if $p < q$, always grant if $p > q$. If $p = q$, the Gatekeeper can commit to any $(\alpha_R, \alpha_{\neg R})$ that keeps the Requesters indifferent. The continuum of equilibria remains. Commitment does not eliminate the multiplicity at the critical threshold. The detailed proof is left to the Appendix.

Finally, it is useful to remark upon the appropriateness of a model that assumes one type of player might be Insane. Note that nothing will change in terms of the results if Insane types were non-strategic and essentially never choose to request not to fly because they want to be able to fly. In this case, while the Gatekeeper would infer that all the Insane never request not to fly (that is, $\beta_I = 1$, it is still optimal for the Gatekeeper to deny all requests to not fly so long as p is sufficiently low as characterised in Proposition 1. Thus, involving potentially irrational, Insane players does not change the outcomes thus far. Nonetheless, a solid takeaway from this analysis is that in high-stakes settings where sanity matters, investment in proper mental health evaluation is worthwhile.[4]

4 Labour Inexperience as a Catch-22

The previous sections presented a formal model of a Catch-22 situation, focusing on the logical paradox where access to a desired state (option A) is only granted to those who do not request it, thereby making it impossible to achieve for those who desire it. It was shown there that the equilibrium outcome, while observationally equivalent to a Catch-22 situation, was not supported by the inference associated with a Catch-22 situation. We now adapt this framework to a labour market setting, where the "experience" required by employers before

---

[4] Unlike apparently authoring papers like this where surely one might reasonably question the sanity of the author in producing this research.

hiring for a good job cannot be obtained without first being hired. This potentially creates a Catch-22 scenario: "You can't get a job without experience, but you can't get experience without a job."[5] Note, however, that this Catch-22 situation is simply a rule and does not have an underlying inference component. Thus, we explore the conditions under which the rule emerges in equilibrium.

In this labour market context, consider a fixed pool of workers of total mass N. A proportion p of these workers is inherently of high potential (H-type), while the remaining $(1-p)$ are low potential (L-type). Workers start as inexperienced, and no external signalling or credentialing is available. The only way for a worker to become "experienced" (and thus reveal or upgrade to their full productivity potential) is through employment. After one period of employment, an inexperienced worker's true type is revealed, or their productivity improves if the model assumes automatic upgrading. Either way, "experience" here means having been employed for a period, allowing the firm to identify or realise the worker's higher productivity if they are H-type.

The firm faces a downward-sloping labour demand function. Let E denote total employment. The marginal product of an experienced H-type worker is $H(E)$, and that of an experienced L-type worker is $L(E)$, with $H'(E) < 0$ and $L'(E) < 0$. Initially, workers are all inexperienced, so when the firm considers hiring from the pool of unknown workers, it must pay a wage equal to their expected productivity:

$$w_{\text{inexp}}(E) = pH(E) + (1-p)L(E).$$

If $pH(E)+(1-p)L(E) \leq 0$, the expected contribution of hiring another inexperienced worker is non-positive. In such a scenario, the firm might refuse to hire altogether. Without hiring, no worker gains experience, and the Catch-22 situation materializes: no one can become experienced (and thus no one can prove their higher productivity) because no one is hired. The equilibrium outcome depends critically on the choices made by employers. If employers are willing to take a risk and hire despite low expected productivity, some workers can gain experience and reveal their true productivity over time. If not, the economy remains in a no-hire equilibrium, trapping workers in inexperience forever.

When $pH(E) + (1-p)L(E) \leq 0$ from the start, no hiring occurs, leading to a full-blown Catch-22 with all workers trapped in inexperience. Now, assume conditions are more

---

[5] This thesis is exemplified by the work of Wilson (1987) and Wilson (1996). Wilson examined how the decline of manufacturing jobs and the rise of service-oriented economies led to a scarcity of employment opportunities in urban areas, particularly affecting African American communities. This scarcity created a cycle where individuals could not gain work experience due to the lack of available jobs yet were required to have experience to secure employment. Wilson argued that this cycle contributes to persistent poverty and social isolation in inner-city neighbourhoods.

favourable at the outset so that at $E = 0$:

$$pH(0) + (1 - p)L(0) > 0.$$

This means that initially, hiring an inexperienced worker is profitable. As a result, the firm begins by hiring some workers, granting them experience and revealing (or upgrading) their potential.

However, both $H(E)$ and $L(E)$ are downward-sloping in E, reflecting diminishing marginal returns as employment grows. As the firm adds more workers, the marginal productivity of both experienced and inexperienced hires declines. Eventually, the firm may reach an employment level E^* where:

$$pH(E^*) + (1 - p)L(E^*) = 0.$$

At this point, hiring another inexperienced worker yields zero expected productivity, leaving the firm indifferent or unwilling to hire further. Workers beyond E^* remain unemployed and inexperienced. They are stuck in a Catch-22: they need to be hired to gain experience and increase their productivity to $H(E)$ if they are H-type, but the firm will not hire them because doing so no longer yields positive returns.

In this scenario, the Catch-22 manifests as an equilibrium outcome affecting some but not all workers. Not all workers are blocked from employment; instead, a subset E^* is hired and becomes experienced, while the remainder $N - E^*$ is left in a state of permanent inexperience and unemployment. Those remaining workers face a Catch-22:

- They cannot become experienced without being hired.
- The firm will not hire them because at equilibrium $pH(E^*) + (1 - p)L(E^*) = 0$, hiring another inexperienced worker yields no net benefit.

This equilibrium differs from the earlier scenario, where no one was hired at all. Here, some are lifted out of inexperience, but the rest remain trapped. The firm's initial willingness to hire disappears once it saturates employment at E^*, where returns from additional inexperienced hires vanish.

These results are summarised in the following:

Proposition 2 *Suppose* $pH(0)+(1-p)L(0) > 0$ *and that* $H(E)$, $L(E)$ *are strictly decreasing with* E, *with* $H(E) > L(E)$ *for all* E. *Then there exists an equilibrium employment level* E^* *such that:*

$$pH(E^*) + (1 - p)L(E^*) = 0,$$

and the firm hires exactly E^* workers, leaving $N - E^*$ workers permanently unemployed and inexperienced. The unemployed workers are stuck in a Catch-22, as they cannot gain experience without employment, and the firm will not hire them at zero expected marginal productivity.

Unlike the initial scenario where no one was hired (full Catch-22), here partial sorting occurs. A fraction E^* is extracted from the pool and becomes experienced, but the equilibrium stops short of employing everyone. The remaining $N - E^*$ are effectively locked out, unable to break into employment and upgrade their productivity. This final outcome still represents a Catch-22 equilibrium for the subset of workers left behind. They cannot become experienced because they cannot get hired, and the firm's equilibrium decision prevents any further expansion. While this might be a Catch-22 situation, it is basically a common result in labour economics; see, for example, Rosen (1981) and MacDonald (1982).

5 Conclusion

This paper has explored the logic, structure, and conditions of Catch-22 situations in a game-theoretic context. By examining two distinct yet parallel scenarios—a stylized version inspired by Joseph Heller's original Catch-22 and a labour market setting where workers cannot gain experience without first having a job—we have shown how such paradoxical outcomes can emerge as stable equilibria in interactive decision-making environments.

In both cases, a single underlying pattern emerges: individuals (Seekers) desire access to some valuable state or resource (e.g., relief from duties, a high-paying job), but the conditions imposed by an authority or gatekeeper (e.g., an officer, an employer) create a logical or strategic deadlock. The Seeker must possess a certain attribute or meet a prerequisite that can only be acquired after obtaining the very outcome they are seeking. As a result, the Seeker is trapped in a loop where attempting to become eligible invalidates their eligibility or where the inability to initiate a required change prevents them from ever qualifying.

Our models highlight that these Catch-22 outcomes do not arise from mere logical trickery but rather from the equilibrium interplay of strategies, incentives, and information constraints. In the literary scenario, a soldier's rational request to avoid a dangerous mission signals sanity, thereby disqualifying him from relief. In the labour market scenario, inexperienced workers remain unemployable if initial conditions fail to justify hiring them, blocking their only avenue to gain the experience needed to become employable. In both, the Catch-22 emerges as a stable configuration of actions and beliefs: the gatekeeper's best response leaves the Seeker stuck, while no profitable deviation exists for the Seeker to escape the predicament.

Our analysis suggests a general structure for identifying and characterizing Catch-22 situations in strategic settings:

1. Players: There are at least two types of players—Seekers who desire a beneficial outcome and Gatekeepers who control access to that outcome. The Seeker cannot autonomously achieve the desired state; the Gatekeeper must grant it or at least provide the necessary conditions.

2. Actions and Information: The Seeker takes actions (e.g., requesting relief, applying for a job) to demonstrate eligibility, while the Gatekeeper responds by accepting or rejecting based on observed signals, beliefs, and payoff structures. Typically, the Gatekeeper's decision rests on information that can only be credibly revealed or produced if the Seeker already had the outcome. This creates a circular dependency.

3. Conditions for a Catch-22: A Catch-22 arises under certain equilibrium conditions:

   - The Gatekeeper sets a condition that is logically or endogenously impossible for the Seeker to meet without already having what they seek.
   - There is no off-equilibrium path or profitable deviation that could break the cycle. Parameters (such as probabilities, costs, marginal products, or payoff differentials) push the equilibrium toward a no-improvement state for the Seeker.
   - Asymmetric information or incomplete verifiability ensures the Seeker cannot credibly demonstrate their eligibility without being granted the outcome first.

4. Why and When They Arise: Such paradoxes often occur in environments with:

   - Asymmetric Information: The Gatekeeper cannot directly observe the Seeker's true state, leading to screening rules that inadvertently exclude everyone.
   - Strict Prerequisites: The qualification the Gatekeeper demands is locked behind the very gate they guard.
   - Unfavourable Parameter Ranges: The model's parameters (e.g., costs, payoffs, productivity levels) may fail to sustain any equilibrium in which the Seeker escapes the loop. In the labour market model, for instance, if hiring inexperienced workers is never profitable beyond a certain point, a nontrivial fraction of workers remains inexperienced forever.

The two scenarios we presented illustrate how different settings can produce structurally similar Catch-22 outcomes. In the original Catch-22 scenario, the condition "you must be insane to be relieved" and the logic that "requesting relief proves sanity" ensures no sane individual can qualify, trapping them in perpetual duty. In the labour market scenario, needing the experience to be hired and needing to be hired to gain experience creates an analogous loop. Initially, positive conditions may allow some workers to escape the trap, but diminishing returns and equilibrium constraints can still leave a subset of workers permanently locked out. In both cases, the paradox does not result from whimsical or ad hoc rules; rather, it emerges naturally from rational decision-making and equilibrium reasoning. The Gatekeeper chooses strategies that maximize their payoff given their beliefs, while the Seeker optimizes their actions within the constraints set by the Gatekeeper's equilibrium strategies. The paradoxical state is stable precisely because it aligns with all players' equilibrium incentives and the structural constraints of the game.

Understanding Catch-22 situations in formal terms opens avenues for both theoretical refinement and practical insight. Identifying the parameters or policy changes that could eliminate or mitigate the Catch-22 might be of direct interest. In the labour market case, for example, subsidies for hiring inexperienced workers or improved screening technologies that lower the cost of identifying high-potential workers could shift parameters and break the cycle. Similarly, in other domains—bureaucratic processes, credit markets, or admission systems—recognizing the equilibrium logic behind a no-win scenario may inspire interventions that relax prerequisites or provide credible signals, thus enabling Seekers to escape the loop.

In short, the general structure for Catch-22 situations involves a closed loop where the Seeker cannot become qualified without first obtaining the outcome, and the Gatekeeper's best response prevents granting that outcome without prior qualification. They are equilibrium outcomes, not anomalies. Understanding their conditions and character reveals not only why they occur but also hints at how changing underlying parameters might restore a more efficient and equitable equilibrium.

By casting Catch-22s into a strategic and equilibrium-centric framework, we gain a deeper understanding of how such paradoxes arise and persist. They are not mere narrative curiosities; they represent stable outcomes embedded in the logic of incentives, information, and constraints. Recognizing these structural elements paves the way for addressing them: by altering rules, improving information flows, or adjusting payoffs, it may be possible to shift from a no-improvement Catch-22 equilibrium to one in which Seekers can genuinely achieve the outcome they desire. In doing so, the analysis offered here contributes not only to economic theory but also to the understanding of institutional and organisational design where Catch-22-like predicaments pervade.

A Appendix: Proofs

A.1 Proof of Proposition 1

Step 1: Gatekeeper's Indifference Condition. The Gatekeeper is indifferent after seeing R if:

$$E[G|R] = E[D|R].$$

Since:

$$E[G|R] = P(I|R)g_I(A) + (1 - P(I|R))g_S(A),$$

$$E[D|R] = P(I|R)g_I(B) + (1 - P(I|R))g_S(B).$$

Set $E[G|R] = E[D|R]$:

$$P(I|R)g_I(A) + (1 - P(I|R))g_S(A) = P(I|R)g_I(B) + (1 - P(I|R))g_S(B).$$

Rearrange terms:

$$P(I|R)(g_I(A) - g_I(B)) + (1 - P(I|R))(g_S(A) - g_S(B)) = 0.$$

Substitute $\Delta_I = g_I(A) - g_I(B) > 0$, $\Delta_S = g_S(A) - g_S(B) < 0$:

$$P(I|R)\Delta_I + (1 - P(I|R))\Delta_S = 0.$$

Solve for $P(I|R)$:

$$P(I|R)(\Delta_I - \Delta_S) = -\Delta_S \implies P(I|R) = \frac{-\Delta_S}{\Delta_I - \Delta_S} = q.$$

A similar derivation shows $P(I|\neg R) = q$ for indifference after $\neg R$.

Step 2: Requester Indifference Conditions. For the Sane type:

$$r_S(R) = \alpha_R r_S(A) + (1 - \alpha_R)r_S(B), \quad r_S(\neg R) = \alpha_{\neg R} r_S(A) + (1 - \alpha_{\neg R})r_S(B).$$

If $r_S(R) = r_S(\neg R)$:

$$\alpha_R r_S(A) + (1 - \alpha_R)r_S(B) = \alpha_{\neg R} r_S(A) + (1 - \alpha_{\neg R})r_S(B).$$

Expand:

$$\alpha_R r_S(A) + r_S(B) - \alpha_R r_S(B) = \alpha_{\neg R} r_S(A) + r_S(B) - \alpha_{\neg R} r_S(B).$$

Subtract $r_S(B)$:

$$\alpha_R(r_S(A) - r_S(B)) = \alpha_{\neg R}(r_S(A) - r_S(B)).$$

Since $r_S(A) - r_S(B) > 0$, we have $\alpha_R = \alpha_{\neg R}$. Similarly, for Insane:

$$r_I(R) = \alpha_R r_I(A) + (1 - \alpha_R) r_I(B), \quad r_I(\neg R) = \alpha_{\neg R} r_I(A) + (1 - \alpha_{\neg R}) r_I(B).$$

$r_I(R) = r_I(\neg R)$ gives, using the same algebraic procedure, $\alpha_R = \alpha_{\neg R}$. Both conditions yield $\alpha_R = \alpha_{\neg R} = \alpha$.

Step 3: Solve Under Different Cases of p vs. q. Case (i) $p < q$: If the Gatekeeper wants to maximise their payoff, consider a deviation from always denying. If the Gatekeeper tries $\alpha > 0$, since $p < q$, the marginal effect of granting increases the likelihood of a worse scenario. In fact, the expected payoff decreases with α. Thus optimal $\alpha = 0$. With always deny, Requesters get B regardless. Neither type can improve by switching messages, as Gatekeeper never changes action. Thus, in equilibrium, the Gatekeeper always denies, $(\alpha_R, \alpha_{\neg R}) = (0, 0)$, and (β_S, β_I) arbitrary.

Case (ii) $p > q$: By symmetry, now granting is beneficial. If Gatekeeper mixes, any reduction in granting lowers their payoff. So $\alpha = 1$; or always grant. Requesters always get A, and so have no incentive to deviate. Thus, in equilibrium, $(\alpha_R, \alpha_{\neg R}) = (1, 1)$, (β_S, β_I) arbitrary.

Case (iii) $p = q$: Set $p = q$. For Gatekeeper to remain indifferent, we need $P(I|R) = p$ and $P(I|\neg R) = p$. From $P(I|R) = \frac{p\beta_I}{p\beta_I + (1-p)\beta_S} = p$, solve:

$$p = \frac{p\beta_I}{p\beta_I + (1-p)\beta_S} \implies p\beta_I + (1-p)\beta_S = p\beta_I/p \implies \beta_S = \beta_I.$$

Similarly, from $P(I|\neg R)$, we get no contradiction; $\beta_S = \beta_I$ suffices.

With $\beta_S = \beta_I$, $P(I|R) = P(I|\neg R) = p$, and the Gatekeeper is indifferent. Any α works. Requesters are indifferent to messages. Thus, there is a continuum of equilibria:

$$\beta_S = \beta_I \in [0, 1], \quad \alpha_R = \alpha_{\neg R} = \alpha \in [0, 1].$$

This proves the proposition.

A.2 Proof of Equilibria When the Gatekeeper Commits First

In this subsection, we consider the alternative timing where the Gatekeeper announces and commits to a policy before the Requester chooses their message. Specifically, the timing is:

1. Nature determines the Requester's type: S with probability $1-p$ or I with probability p.
2. The Gatekeeper publicly commits to a policy $(\alpha_R, \alpha_{\neg R})$, where α_R is the probability of granting after observing R, and $\alpha_{\neg R}$ is the probability of granting after observing $\neg R$.
3. After seeing the Gatekeeper's commitment, the Requester chooses R or $\neg R$ to maximize their own expected payoff given their type.

We consider a Subgame Perfect Equilibrium (SPE). The Gatekeeper chooses $(\alpha_R, \alpha_{\neg R})$ first. Given that policy, the Requester, knowing their type, chooses R or $\neg R$. Because the Requester moves last, they pick the action that maximizes their expected utility, taking $(\alpha_R, \alpha_{\neg R})$ as given. The Gatekeeper anticipates this response and chooses an optimal policy.

Requester's Best Response Given $(\alpha_R, \alpha_{\neg R})$

For a given Gatekeeper policy: - If the Requester is Sane, their expected payoffs are:

$$r_S(R) = \alpha_R r_S(A) + (1-\alpha_R) r_S(B), \quad r_S(\neg R) = \alpha_{\neg R} r_S(A) + (1-\alpha_{\neg R}) r_S(B).$$

Since $r_S(A) > r_S(B)$, the Sane type compares $r_S(R)$ and $r_S(\neg R)$:

$$r_S(R) - r_S(\neg R) = (\alpha_R - \alpha_{\neg R})(r_S(A) - r_S(B)).$$

- If the Requester is Insane:

$$r_I(R) = \alpha_R r_I(A) + (1-\alpha_R) r_I(B), \quad r_I(\neg R) = \alpha_{\neg R} r_I(A) + (1-\alpha_{\neg R}) r_I(B).$$

With $r_I(B) > r_I(A)$, we have:

$$r_I(R) - r_I(\neg R) = (\alpha_R - \alpha_{\neg R})(r_I(A) - r_I(B)).$$

Note $(r_I(A) - r_I(B)) < 0$, so if $\alpha_R > \alpha_{\neg R}$, the Insane type is more likely to get A (unwanted), decreasing their utility relative to $\neg R$; if $\alpha_R < \alpha_{\neg R}$, the reverse is true.

Gatekeeper's Anticipation

The Gatekeeper knows that after commitment, the Requester chooses the message to maximize their payoff. Consider what happens under different configurations of $(\alpha_R, \alpha_{\neg R})$:

1. If $\alpha_R > \alpha_{\neg R}$: the Sane type prefers R since that yields a higher probability of A, their preferred outcome. The Insane type prefers $\neg R$ since $\alpha_R > \alpha_{\neg R}$ makes R more likely to yield A, which they dislike. Thus, the Insane type avoids R. In this scenario, the messages separate types: Sane chooses R, Insane chooses $\neg R$.

2. If $\alpha_R < \alpha_{\neg R}$ the Sane type prefers $\neg R$ (since $\neg R$ has a higher probability of A now). The Insane type prefers R (since R now has a lower probability of A and thus a higher chance of B, their preferred outcome). Here, the messages also separate types but in the opposite manner: Sane chooses $\neg R$, Insane chooses R.

3. If $\alpha_R = \alpha_{\neg R}$ both Sane and Insane are indifferent between R and $\neg R$. Hence, the Requesters may mix arbitrarily. There is no separating pattern; either type could choose R or $\neg R$ with any probability.

Given this, the Gatekeeper's payoff depends on how the actions map to outcomes for each type. We use the same logic as in the main text. The Gatekeeper would like Sane types to end up with B and Insane types to end up with A.

Now, consider the three main cases regarding p and q:

Case 1: $p < q$**.** If $p < q$, recall from the main text analysis that when the Gatekeeper attempts to equalize or mix, the equilibrium in the earlier scenario led to always denying. Here, because the Gatekeeper moves first, can they do better? Suppose the Gatekeeper tries some $\alpha > 0$. If α_R and $\alpha_{\neg R}$ differ, the Requester separates: one type gets more A, the other gets more B. The Gatekeeper wants Insane to get A and Sane to get B, but if $p < q$, the population is more likely Sane. If the Gatekeeper chooses $\alpha_R > \alpha_{\neg R}$, Sane chooses R, gets mostly A (which the Gatekeeper does not want). If $\alpha_R < \alpha_{\neg R}$, Sane chooses $\neg R$, gets mostly A anyway (just switching messages), still against the Gatekeeper's preference.

By trying to induce a difference between α_R and $\alpha_{\neg R}$, the Gatekeeper inadvertently ensures that the Sane type will gravitate to the message that yields A. Since $p < q$, the Gatekeeper's expected payoff is worse off granting frequently because the likely Sane Requester gets their preferred outcome A, which the Gatekeeper dislikes.

To avoid this, the Gatekeeper sets $\alpha_R = \alpha_{\neg R} = 0$. This yields no incentive for Sane or Insane to prefer one message over the other; both get B. Insane is happy with B, Sane is not, but cannot do better by switching. The Gatekeeper's expected payoff is:

$$\Pi_{\text{deny}} = pg_I(B) + (1-p)g_S(B).$$

This is exactly the same outcome as in the previously analyzed model where the Requester moved first. The Gatekeeper cannot improve upon always denying. Hence, if $p < q$, the Gatekeeper's optimal committed policy is $(\alpha_R, \alpha_{\neg R}) = (0, 0)$.

Case 2: $p > q$. By symmetry, if the probability of Insanity is high, the Gatekeeper prefers to ensure Insane types get A. If the Gatekeeper tries $\alpha_R < \alpha_{\neg R}$ or vice versa, they end up granting a scenario where Sane might get A. With $p > q$, there are relatively more Insane types, so to ensure high expected payoff, the Gatekeeper sets $\alpha_R = \alpha_{\neg R} = 1$, always granting. In this scenario, both types get A. Insane prefers B, but cannot force the Gatekeeper to deny. Sane is satisfied with A, Insane cannot do better. The Gatekeeper's payoff is:

$$\Pi_{\text{grant}} = pg_I(A) + (1-p)g_S(A).$$

Since $p > q$, this aligns with the Gatekeeper's preference to favor outcomes suitable for Insane. No mixing improves the Gatekeeper's payoff beyond always granting.

Case 3: $p = q$. If $p = q$, the Gatekeeper is exactly indifferent when facing balanced posteriors. By choosing $\alpha_R = \alpha_{\neg R} = \alpha$ for any $\alpha \in [0, 1]$, the Requesters remain indifferent. They can mix in any proportion. No profitable deviation exists for the Gatekeeper by breaking this balance, as any asymmetry would allow the Requesters to self-select into messages that yield an undesired pairing of types and outcomes. Hence, a continuum of equilibria emerges:

$$(\alpha_R, \alpha_{\neg R}) = (\alpha, \alpha), \quad \alpha \in [0, 1],$$

with the Requesters mixing arbitrarily. This exactly mirrors the multiplicity found in the original timing scenario. The Gatekeeper, anticipating the Requesters' best responses, cannot improve by picking a unique pure strategy other than a corner solution, but since $p = q$, corner solutions are no better than mixing solutions.

Conclusion for the Gatekeeper-First Scenario

When the Gatekeeper commits first:

- If $p < q$, the Gatekeeper chooses $(\alpha_R, \alpha_{\neg R}) = (0, 0)$, always denying. Requesters mix arbitrarily, equilibrium outcome matches previously found pure equilibrium.

- If $p > q$, the Gatekeeper chooses $(\alpha_R, \alpha_{\neg R}) = (1, 1)$, always granting. Requesters again mix arbitrarily, equilibrium outcome matches the pure equilibrium of always granting.
- If $p = q$, the Gatekeeper can choose $(\alpha_R, \alpha_{\neg R}) = (\alpha, \alpha)$ for any $\alpha \in [0, 1]$. Requesters mix freely, resulting in a continuum of equilibria that parallel the original scenario.

Thus, committing early does not remove the multiplicity at $p = q$, nor does it lead to a better outcome for the Gatekeeper when $p \neq q$. The results are fully consistent with the analysis conducted under the original timing.

A.3 Proof of Proposition 2

Step 1: Existence of E^*. Start from $E = 0$ where $pH(0) + (1-p)L(0) > 0$. By continuity and the fact that $H(E)$ and $L(E)$ decrease with E, there must be some largest $E > 0$ at which $pH(E) + (1-p)L(E) = 0$. Define E^* as that point:

$$pH(E^*) + (1-p)L(E^*) = 0.$$

Such an E^* exists because at $E = 0$ the value is positive, and as $E \to \infty$, $H(E), L(E)$ become very low (possibly negative or approaching zero), ensuring that eventually the expected productivity crosses zero.

Step 2: Optimality of Hiring Up to E^*. For $E < E^*$, we have $pH(E)+(1-p)L(E) > 0$, so hiring another inexperienced worker is profitable. The firm continues hiring until E reaches E^*. At E^*, the next marginal hire yields zero expected return, so the firm stops hiring further to avoid negative gains.

Step 3: Catch-22 for $N - E^*$ Workers. At equilibrium E^*, there are $N - E^*$ workers who remain untested and inexperienced. Their potential productivity could be high if they are H-type, but they never gain the chance to demonstrate this because the firm refuses to hire more workers once the marginal expected productivity hits zero. These workers are stuck in the situation that defines Catch-22: (i) they need a job to gain experience and reveal or achieve higher productivity and (ii) the firm will not hire them now that the expected return is zero at the margin.

References

Brams, Steven J, “To mobilize or not to mobilize: Catch-22s in international crises,” *International Studies Quarterly*, 1999, *43* (4), 621–640.

Brams, Steven J. and Christopher B. Jones, “Catch-22 and King-of-the-Mountain Games: Cycling, Frustration, and Power,” *Rationality and Society*, 1999, *11* (2), 139–167.

Carmichael, Fiona, “Catch-22 in a Signalling Game,” *Journal of Institutional and Theoretical Economics*, 2002, *158* (3), 375–392.

Colgan, Charles S. and Gary Quinlin, “The Catch-22 of Congestion Pricing,” *Transportation Quarterly*, 1997, *51* (2), 117–133.

Crawford, Vincent P. and Joel Sobel, “Strategic Information Transmission,” *Econometrica*, 1982, *50* (6), 1431–1451.

de Rouffignac, A., “Supply vs. Demand: Catch-22 for US Gas,” *Energy Economist*, January 2000, (219), 6–9.

Flood, P. C. and B. Toner, “La Catch 22?,” *British Journal of Industrial Relations*, 1997, *35* (1), 123–124.

Goldstein, Laurence, “The Barber, Russell’s Paradox, Catch-22, God, Contradiction, and More,” in Graham Priest, J. C. Beall, and Bradley P. Armour-Garb, eds., *The Law of Non-Contradiction: New Philosophical Essays*, New York: Oxford University Press, 2004, pp. 295–313.

Heller, Joseph, *Catch-22*, New York: Simon & Schuster, 1961.

MacDonald, Glenn M., “Information in Production,” *Econometrica*, 1982, *50* (5), 1143–1162.

Rosen, Sherwin, “An Equilibrium Approach to the Labor Market,” *Econometrica*, 1981, *49* (3), 565–584.

Stein, M. K., “The Catch-22 of Retirement,” *Journal of Financial Planning*, 1999, *12*, 28–30.

Wilson, William Julius, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, University of Chicago Press, 1987.

_ , *When Work Disappears: The World of the New Urban Poor*, Vintage Books, 1996.