# Reciprocal Preference and Expectations in International Agreements

Suha Kim[1] Doruk Iris[2]

[1]Ohio State University

[2]Sogang University

ESA Charlotte 2023

# Motivation

Ample evidence that **individuals** have reciprocal preferences: the desire to be kind towards kind and unkind towards unkind behavior.

- Andreoni, 1988; Camerer, 2003; Charness and Rabin, 2002; Croson, 2007; Dohmen et al., 2009; Falk et al., 2003, 2008; Falk and Fischbacher, 2006; Palfrey and Prisbrey, 1997

Coalition formations to provide public goods (e.g., tackling environmental problems, eradication of epidemics, etc.)

How does this reciprocal preference affect a coalition to provide public goods? Formalize in theory?

# Main Questions

If reciprocal preferences exist for countries to some degree,

- In what ways do these **reciprocal preferences** affect international agreements to provide global public goods?
- How do countries' **expectations** towards others (i.e., which behaviors are perceived as kind or unkind) affect such agreements?
- How would a country behave if it strategically uses its expectations towards others?

➔ To address, countries face

- public goods dilemma in a coalition formation game (*Barrett (1994)*)
- Extend and incorporate reciprocal preferences model(*Rabin (1993)*)

# Related Literature

Compared to the existing literature, this paper:

- Continuous choice of contribution:
  - Countries not only decide to participate or not (0 or 1), (Nyborg (2017,JEEM), Buchholz et al. (2018, JEBO))
  - but also how much effort to exert (Lange and Vogt, 2003, JPubE): Inequality-aversion
- Extend *fair effort threshold*, capturing countries' expectations towards others:
  - How the high/low/moderate expectations toward the other's kindness can affect the reciprocal behavior / coalition

# Model

We follow Nyborg (2018) closely for the sake of transparency.

- $N$ (identical) countries, each decides an effort level $q_i \in [0, 1]$.
- A reciprocal country $i$'s utility:

$$U_i = \Pi_i + \alpha R_i$$

1. Material payoff $\Pi_i$
   - $\Pi_i = bQ - \frac{c}{2}q_i^2$ where $Q = \sum_i q_i$, $b, c > 0$
   - Public goods game
     $\begin{cases} b < c & \rightarrow \text{Individual incentive to free ride} \\ c < Nb & \rightarrow \text{Everyone exerting full effort is the social optimum} \end{cases}$
2. Social payoff (Reciprocal payoff) $\alpha R_i$
   - (Rabin (1993))
     $R_i$ consists of $\underbrace{\text{kindness functions}}_{\substack{f_{ij}\text{: payoffs that} \\ i \text{ can secure to } j}}$ & $\underbrace{\text{Equitable payoff}}_{\substack{\eta\text{: threshold expected payoffs} \\ \text{that is believed by } i \text{ to be kind}}}$

# Model

Equitable payoff and $f_{ij}$

- Reciprocal payoff:

$$R_i = \frac{1}{N-1}\left[\sum_{j\neq i}\tilde{f}_{ji} + \sum_{j\neq i} f_{ij}\tilde{f}_{ji}\right]$$
$$= \left(\frac{\tilde{Q}_{-i}}{N-1} - \eta\right) + (q_i - \eta)\left(\frac{\tilde{Q}_{-i}}{N-1} - \eta\right)$$

- Utility function:

$$U_i = \Pi_i + \alpha R_i$$
$$= \left(b(\tilde{Q}_{-i} + q_i) - \frac{c}{2}{q_i}^2\right) + \alpha\left(\frac{\tilde{Q}_{-i}}{N-1} - \eta\right)(1 + q_i - \eta)$$

# Model

- Coalition formation game structure to provide public goods
- Single coalition

Coalition formation in 3 Stages:

**Stage 1.** Every country $i$ decides simultaneously and independently whether to sign or not to sign the treaty. (Let $k$ denote the number of signatories.)

**Stage 2.** Signatories decide their strategies collectively, maximizing their joint payoff.

**Stage 3.** Non-signatories choose their strategies non-cooperatively.

# 1) Non-cooperative game - Benchmark

✱ For a reference, what happens in the extreme cases?

✱ What is the impact of the reciprocity in Non-cooperative effort decisions?

## Proposition 1

- Non-cooperative Reciprocal / Non-cooperative Self-interested contributions:

$$q_{NC}^{R} = \frac{b-\alpha\eta}{c-\alpha}, \qquad q_{NC}^{S} = \frac{b}{c}$$

- $q_{NC}^{R} \leq q_{NC}^{S}$ iff $\eta \geq q_{NC}^{S}$
  - Expectation toward another countries' kindness: higher than self-interested Nash
  - Higher bar to perceive kindness $\rightarrow$ less effort when reciprocity introduced
- $q_{NC}^{R} \leq 1$ iff $\alpha \leq \frac{c-b}{1-\eta}$, $q_{NC}^{R} = 1$ iff $\alpha \geq \frac{c-b}{1-\eta}$
  - With strong enough reciprocity, countries give full effort even non-cooperately. (* Nyborg (2018), Buchholz et al. (2018))

# 2) Partial Cooperation

✱ What about in less extreme cases, where $k$ out of $N$ countries participate? (Backward Induction in stages)

- Again, for sufficiently high reciprocal concerns $\alpha$ and fair effort threshold $\eta$ not so high, both signatories and non-signatories exert 1. (Corner Solution, Corollary)
- From now on, focus on low reciprocity: $\alpha \leq \frac{c-b}{1-\eta}$
  - $\alpha$ becomes sufficiently small if stakes are high (Rabin (1993))

# 2) Partial Cooperation

Non-signatories & Signatories effort levels

## Proposition 2

- Signatory and Non-signatory efforts under self-interest:

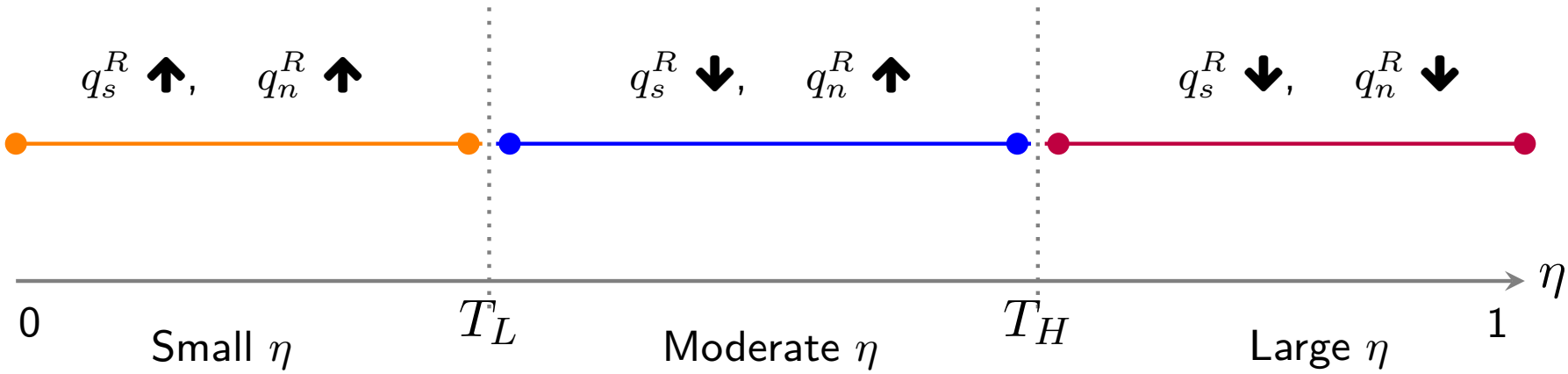
$$q_s^S = bk/c, \qquad q_n^S = b/c$$

- Under reciprocal preferences, non-signatories do not have dominant strategy anymore
- Signatories' efforts positively and linearly affect non-signatories' efforts (Direct impact) : $\frac{dq_n^R}{dq_s^R} > 0$
  - Consistent with the results in Leader-follower public goods experiments
- Knowing that non-signatories' positive respond, signatories have an additional incentive to increase their effort level (2nd degree impact)

# 2) Partial Cooperation

✱ How does the reciprocity ($\alpha$) impact on the effort levels?

Examining the impact of $\alpha$ by taking derivatives of the effort levels ($q_s^R$ and $q_n^R$) around $\alpha = 0$, we can see this:



# 2) Partial Cooperation - Stability

✱ How many countries participate in a stable treaty?

**Definition (Stable Coalition Size)**

A coalition of size $k$ is stable if $U_s(k) - U_n(k-1) \geq 0$ (i.e., internal stability) and $U_s(k+1) - U_n(k) \leq 0$ (i.e., external stability.)

- Under the standard preferences ($\alpha = 0$),
  our model gives the stable coalition size 3.
  ($U_s(k) = U_n(k-1)$ at $k = 3$)
- Under reciprocal preferences ($\alpha > 0$),
  we showed that the stable coalition size uniquely exists and it is either 2 or 3.

# 2) Partial Cooperation - Stability

**Proposition 4**

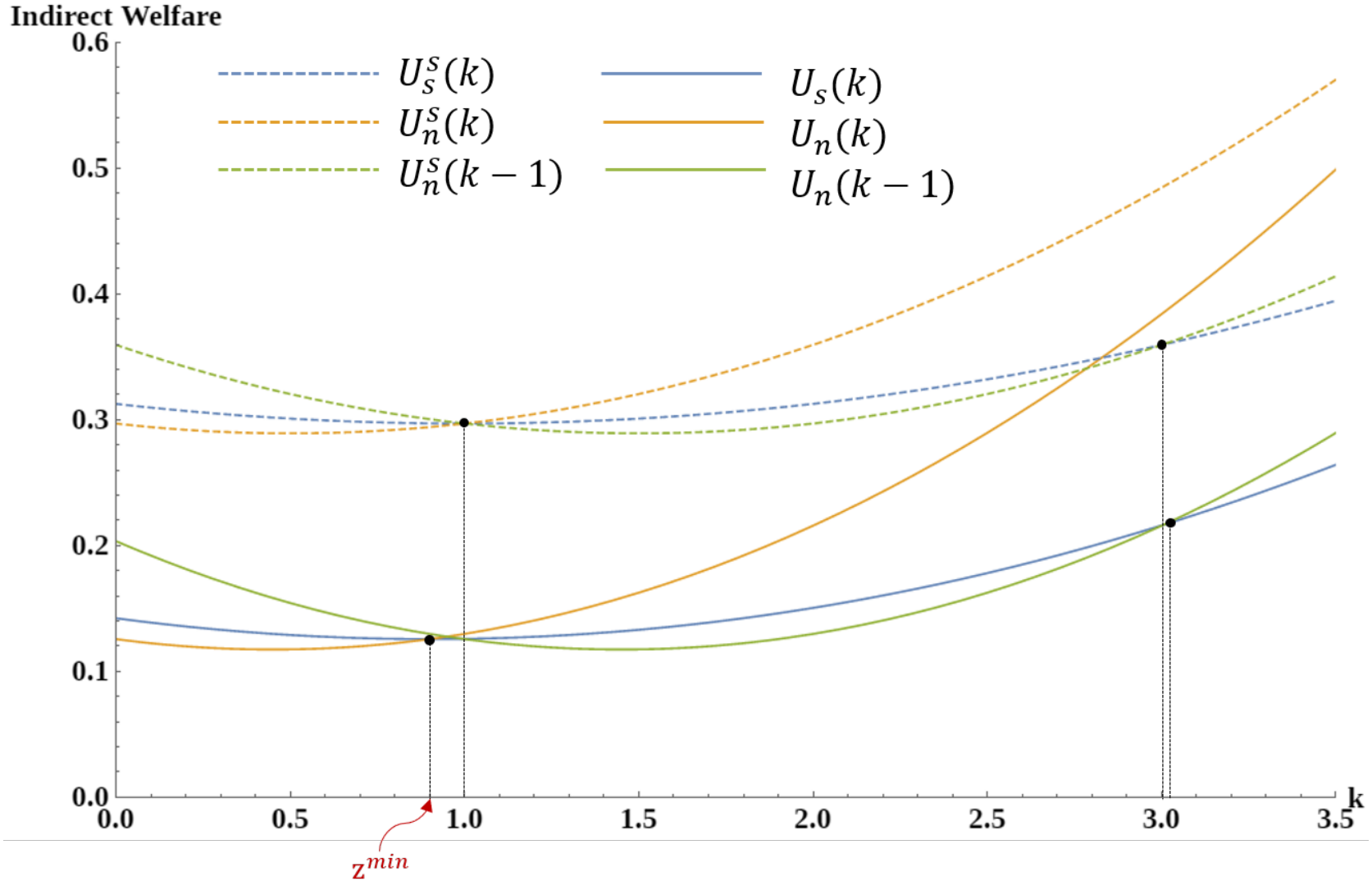
If $\alpha$ is sufficiently small, then the stable treaty size can be $k^* \in \{2, 3\}$ for $N \geq 3$. Introduction of such reciprocal concerns with some moderate expectations, i.e., $\eta \in [T'_L, T_L]$, suffices for the stable coalition size to shrink from 3 to 2.

Given the size of the stable treaty is always 3 under self-interested preferences, reciprocal preferences can decrease the size of the stable treaty but cannot increase it!

- This result is particularly important:
  - The binary choice models (Nyborg (2018), Buchholz (2018)) find that full cooperation is possible if the reciprocal concern ($\alpha$) is strong enough.

# 2) Partial Cooperation - Shrinking Stable Size

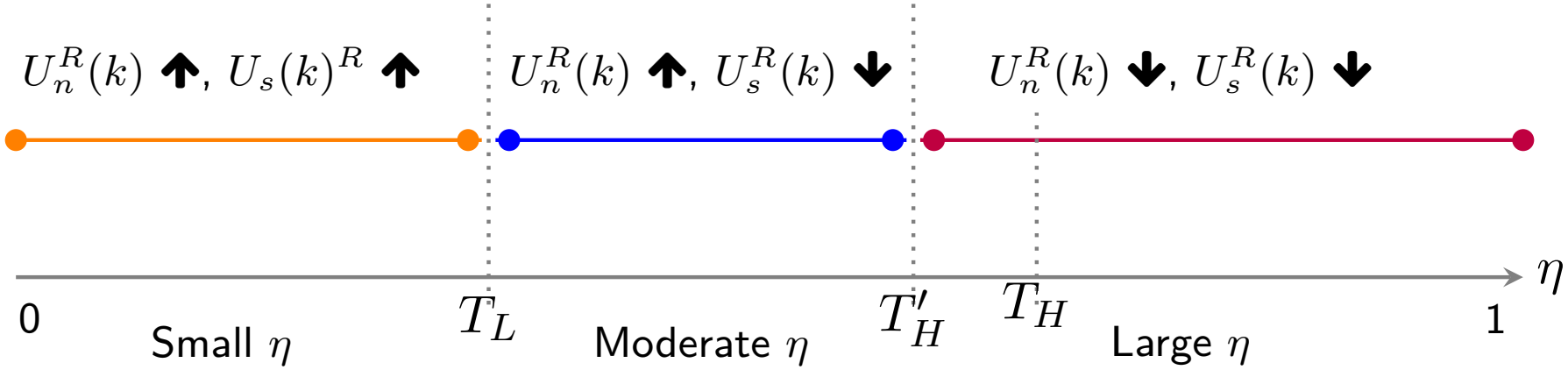
✱ When does the stable size of coalition shrink to 2?



# 2) Partial Cooperation - Shrinking Stable Size

✱ When does the stable size of coalition shrink to 2?

Examining the impact of $\alpha$ by taking derivatives of the signatories' / non-signatories' indirect welfare functions ($U_n^R(k)$ and $U_s(k)^R$) around $\alpha = 0$, we can see this:




The stable coalition size can shrink with moderate expectations

# 2) Partial Cooperation - High Reciprocity

- Might be due to assuming sufficiently small $\alpha$:
  - Investigate whether there are any feasible parameter values that produce interior solutions and yield a stable grand coalition
  - A condition $U_s(N) \geq U_n(N-1)$
  - Numeric example:
    e.g., $b = 0.0307, c = 1.01, N = 10, \alpha = 0.5, \eta = 0.8$
    ➜ grand coalition to be stable with the effort level $q_s^R = 0.9$.

**Result 5**

If $\alpha$ takes sufficiently high values, then the grand coalition with interior effort levels can be stable.

# Conclusion: Stable Treaty

- The grand coalition is stable for sufficiently high reciprocal concerns; Nyborg's (2017) main result is robust
- However, if the solution for effort levels is interior, the impact of reciprocal concerns is limited
- Stable coalition size can be $\{2, 3\}$
- A sufficiency condition: Stable coalition size would shrink to $2$ if countries have moderate expectations

# Conclusion: Efforts

- Introduction of reciprocal preference $\Longrightarrow$ non-signatories do not have dominant strategy anymore
- Signatories' efforts $\uparrow \Longrightarrow$ Non-signatories' effort $\uparrow$
- Knowing this, signatories have an additional incentive to increase efforts
  - EU's leadership in climate action

- Countries' expectations (low, moderate, or high) plays a significant role

# Thank you!

kim.8316@osu.edu

# Extension - Strategic use of expectations

✱ How would a country with reciprocal preferences strategically use its expectations ($\eta$) towards others?

**Stage 1.** Each country $i$ declares $\eta_i$
**Stage 2.** Each country $i$ determines its non-cooperative efforts

**Setup:**

- $N-1$ countries non-strategically declare their truthful $\eta_T$, while a single country unilaterally and strategically announces its $\eta_S$
- Countries believe others' fair threshold expectations to be true.

**Assumptions:**

1. Self-fulfilling prophecy: The strategic country $S$ also believes its announcement $\eta_S$ to be true.
2. Self-awareness: The strategic country $S$ is aware of its true parameter value $\eta_T$, but strategically declare $\eta'_S$.

# Extension - Strategic use of expectations

Under the self-fulfilling prophecy, country $S$ maximize the following utility function ($U_T$ defined similarly):

$$U_S = b((N-1)q_T + q_S) - \frac{c}{2}q_S^2 + \alpha(q_T - \eta_S)(1 + q_S - \eta_T).$$

$$q_S(\eta_T, \eta_S) = \frac{b(c(N-1)+\alpha) - \alpha^2(N-1)\eta_T - ((N-1)(c-\alpha)+\alpha^2)\eta_S}{(c-\alpha)(c(N-1)+\alpha}$$

$$q_T(\eta_T, \eta_S) = \frac{b(c(N-1)+\alpha) - \alpha(c(N-1)\eta_T + \alpha\eta_S)}{(c-\alpha)(c(N-1)+\alpha}$$

3 forces that determine the strategically chosen $\eta_S$:

- Both $\eta_T$ and $\eta_S$ enter negatively to the effort levels
- Each type's own $\eta$ decreases their own effort level faster than the other type's $\eta$.
- Utility: an increase in $\eta_S$ also decreases utility since country $S$ perceives others being more unkind or less kind

# Extension - Strategic use of expectations

Two intuitive effects on Country $S$'s strategic use of $\eta$:

- By strategically setting $\eta_S > \eta_T$, country $S$ could find an excuse to lower its effort (and lower cost to bear), while a higher $\eta_S$ only marginally lowers other countries' efforts.
- It also has incentive to lower $\eta_S$ since it helps perceiving others kinder.

**Result: Self-fulfilling prophecy**

- If $\eta_T$ is very high, then for some parameters $b, c, \alpha$ with small $b/c$, country $S$ can set $\eta_S < \eta_T$
- Otherwise, country $S$ sets $\eta_S > \eta_T$
- Under all conditions, country $S$ chooses $\eta_S$ such that it perceives others as unkind

# Extension - Strategic use of expectations

Self-awareness:

- Different than the analysis under the self-fulfilling prophecy, $\eta'_S$ has no direct impact on the utility, because country $S$ knows that its true expectations is $\eta_T$
- Since there is no force decreasing $\eta'_S$, country $S$ use $\eta'_S$ always to find excuse to lower its effort level by setting $\eta'_S \geq \eta_S$

**Result: Self-awareness**

Country $S$ sets $\eta'_S = 1$ and perceives others as unkind.

# Extension - Heterogeneity of Reciprocity

Now assume that preference are given by

$$U_i = \Pi_i + \alpha_i R_i, \tag{1}$$

where $\alpha_i \in \{0, \alpha\}$.
Let $A < N$ be the number of countries with $\alpha_i = \alpha$ and $N - A$ is the number of countries with $\alpha_i = 0$.

## Result

1. Number: There may be no, unique, or multiple stable coalitions.
2. Formation: The stable coalitions can consist of only self-interested countries, only reciprocal countries, or a mixture of the two.
3. Expectations: As $\eta$ increases, more self-interested and less reciprocal countries tend to participate in the coalition.
4. Size: For small $\alpha$, stable coalition size can be $k^* \in \{2, 3\}$. But for sufficiently high $\alpha$, up to $A$ number of reciprocal countries can form a stable coalition.

# Appendix

Equitable payoff and $f_{ij}$

- We give some freedom to the position of $\Pi_{ij}^e$:

$$\Pi_{ij}^e = \eta\Pi_{ij}^{max} + (1-\eta)\Pi_{ij}^{min}, \text{ where } \eta \in (0, 1]$$

  * Rabin (1993), Nyborg (2018): $\Pi_{ij}^e = \frac{1}{2}\Pi_{ij}^{max} + \frac{1}{2}\Pi_{ij}^{min}$

- Then the kindness function $(f_{ij})$ is simplified as:

$$f_{ij} = \frac{\Pi_j(q_i, \tilde{Q}_{-i}) - \Pi_{ij}^e}{\Pi_{ij}^{max} - \Pi_{ij}^{min}} = q_i - \eta \quad \text{(and also, } f_{ji} = q_j - \eta)$$

Back to the slide

# Appendix

Non-signatories and signatories effort levels:

$$q_n^R = \frac{b(N-1) + \alpha(kq_s^R - (N-1)\eta)}{c(N-1) - \alpha(N-k-1)}$$
$$= \frac{dq_n^R}{dq_s^R} q_s^R + \frac{(N-1)(b-\alpha\eta)}{c(N-1) - \alpha(N-k-1)}$$

$$q_s^R = \frac{bk + \alpha\left(\frac{k-1}{N-1}(1-\eta) + \frac{(N-k)(b-\alpha\eta)}{c(N-1)-\alpha(N-k-1)} - \eta\right) + \left(b(N-k) + \alpha\frac{N-k}{N-1}(1-\eta)\right)\frac{dq_n^R}{dq_s^R}}{c - 2\alpha\left(\frac{k-1}{N-1} + \frac{N-k}{N-1}\frac{dq_n^R}{dq_s^R}\right)}$$

Back to the Slide