Over-contribution in discretionary databases

Mike Klaas

klaas@cs.ubc.ca

Faculty of Computer Science

University of British Columbia

Outline

Outline

- Social dilemmas in discretionary databases

Outline

- Social dilemmas in discretionary databases
- Utility model

Outline

- Social dilemmas in discretionary databases
- Utility model
- Agent behaviour in Usenet

Outline

- Social dilemmas in discretionary databases
- Utility model
- Agent behaviour in Usenet
- Group behaviour in Usenet

Outline

- Social dilemmas in discretionary databases
- Utility model
- Agent behaviour in Usenet
- Group behaviour in Usenet
- Value of an item

Outline

- Social dilemmas in discretionary databases
- Utility model
- Agent behaviour in Usenet
- Group behaviour in Usenet
- Value of an item
- Global resource allocation

Outline

- Social dilemmas in discretionary databases
- Utility model
- Agent behaviour in Usenet
- Group behaviour in Usenet
- Value of an item
- Global resource allocation
- Conclusion

Discretionary Databases

Discretionary Databases

- Social dilemma: self interest $\Leftrightarrow$ collective interest

Discretionary Databases

- Social dilemma: self interest $\Leftrightarrow$ collective interest
  - ex. the notorious *free-rider* problem (Sweeney 1973)

Discretionary Databases

- Social dilemma: self interest $\Leftrightarrow$ collective interest
  - ex. the notorious *free-rider* problem (Sweeney 1973)
- Game-theoretic approaches
  - (Golle *et al.* 2001) (Buragohain *et al.* 2003) — emphasizing P2P and free-ridership

Discretionary Databases

- Social dilemma: self interest ⇔ collective interest
  - ex. the notorious *free-rider* problem (Sweeney 1973)
- Game-theoretic approaches
  - (Golle *et al.* 2001) (Buragohain *et al.* 2003) — emphasizing P2P and free-ridership
- Performance models
  - (Fuqua *et al.* 2003) (Mfeldman *et al.* 2003) — still use incentive models

Discretionary Databases

- Social dilemma: self interest $\Leftrightarrow$ collective interest
  - ex. the notorious *free-rider* problem (Sweeney 1973)
- Game-theoretic approaches
  - (Golle *et al.* 2001) (Buragohain *et al.* 2003) — emphasizing P2P and free-ridership
- Performance models
  - (Fuqua *et al.* 2003) (Mfeldman *et al.* 2003) — still use incentive models
- Our goal: analyze over-contribution; reputation

Usenet—Overview

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:
  - free-riding not death-knell

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:
  - free-riding not death-knell
  - shared files always consume resources

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:
  - free-riding not death-knell
  - shared files always consume resources
  - community spirit & feedback

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:
  - free-riding not death-knell
  - shared files always consume resources
  - community spirit & feedback
  - more significant sharing disincentives

Usenet—Overview

- 1970 $\Longrightarrow$ little text traffic
- 2004 $\Longrightarrow$ 1TB/day binary data contributed
- Self-organized distributed file-sharing system
- Differences from P2P:
  - free-riding not death-knell
  - shared files always consume resources
  - community spirit & feedback
  - more significant sharing disincentives
  - over-contribution

Model overview

Model overview

- Bayesian game with $\mathcal{I} = \{a_1, a_2, \ldots, a_n\}$

Model overview

- Bayesian game with $\mathcal{I} = \{a_1, a_2, \ldots, a_n\}$
- Consuming actions
  - eg., downloading a file
  - $u_i^{DN} \triangleq a_i$'s consumption utility

Model overview

- Bayesian game with $\mathcal{I} = \{a_1, a_2, \ldots, a_n\}$
- Consuming actions
  - eg., downloading a file
  - $u_i^{DN} \triangleq a_i$'s consumption utility
- Contributing actions
  - eg., uploading a file
  - $u_i^{UP} \triangleq a_i$'s contributory utility

Model overview

- Bayesian game with $\mathcal{I} = \{a_1, a_2, \ldots, a_n\}$
- Consuming actions
  - eg., downloading a file
  - $u_i^{DN} \triangleq a_i$'s consumption utility
- Contributing actions
  - eg., uploading a file
  - $u_i^{UP} \triangleq a_i$'s contributory utility
- Total utility $u_i \triangleq u_i^{UP} + u_i^{DN}$

Consumption utility (u_i^{DN})

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*
  - *Heterogeneity*

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*
  - *Heterogeneity*
  - *Implicit/Explicit Cost*

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*
  - *Heterogeneity*
  - *Implicit/Explicit Cost*
- Models:
  - Let $\mathcal{Q}$ be all content, $C \subseteq \mathcal{Q}$

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*
  - *Heterogeneity*
  - *Implicit/Explicit Cost*
- Models:
  - Let $\mathcal{Q}$ be all content, $C \subseteq \mathcal{Q}$
  - previously, mostly linear:
    - $u_i^{DN}(C) \propto size(C)$

Consumption utility (u_i^{DN})

- Factors:
  - *Content Retrieved* — possibly capped
  - *Variety*
  - *Heterogeneity*
  - *Implicit/Explicit Cost*
- Models:
  - Let $\mathcal{Q}$ be all content, $C \subseteq \mathcal{Q}$
  - previously, mostly linear:
    - $u_i^{DN}(C) \propto size(C)$
  - problems:
    - doesn't model variety or interest
    - size isn't linear (cd-image 2^{29} vs picture 2^{18})

Consumption utility (framework)

Consumption utility (framework)

- Partition $\mathcal{Q}$ into $\{Q_i\}$, pick $C_i \subseteq Q_i$
  - variety through substitutability

Consumption utility (framework)

- Partition $\mathcal{Q}$ into $\{Q_i\}$, pick $C_i \subseteq Q_i$
  - variety through substitutability
- Interest matrix $\mathbf{W}$
  - w_{ij} is a_i's interest in C_j

Consumption utility (framework)

- Partition $\mathcal{Q}$ into $\{Q_i\}$, pick $C_i \subseteq Q_i$
  - variety through substitutability
- Interest matrix $\mathbf{W}$
  - w_{ij} is a_i's interest in C_j
- Class utility function θ_i
  - ex. $\sqrt{x}$, $\log(1 + x)$

Consumption utility (framework)

- Partition $\mathcal{Q}$ into $\{Q_i\}$, pick $C_i \subseteq Q_i$
  - variety through substitutability
- Interest matrix $\mathbf{W}$
  - w_{ij} is a_i's interest in C_j
- Class utility function θ_i
  - ex. $\sqrt{x}$, $\log(1+x)$
- Cost function $cost_i^{DN}$

Consumption utility (framework)

- Partition $\mathcal{Q}$ into $\{Q_i\}$, pick $C_i \subseteq Q_i$
  - variety through substitutability
- Interest matrix $\mathbf{W}$
  - w_{ij} is a_i's interest in C_j
- Class utility function θ_i
  - ex. $\sqrt{x}$, $\log(1+x)$
- Cost function $cost_i^{DN}$
- Through the combination of these, we can model virtually any cost function
  - $u_i^{DN}(C) = -cost_i^{DN}(C) + \sum_{j=1}^{m} w_{ij}\theta_i(size(C_j))$

Contribution utility (u_i^{UP})

Contribution utility (u_i^{UP})

- Factors:

- Model:

Contribution utility (u_i^{UP})

- Factors:
  - *Inherent Cost*
- Model:
  - cost function $-cost_i^{UP}$

Contribution utility (u_i^{UP})

- Factors:
  - *Inherent Cost*
  - *Inherent Contribution Preference*

- Model:
  - cost function $-cost_i^{UP}$
  - benefit function $gain_i^{UP}$

Contribution utility (u_i^{UP})

- Factors:
  - *Inherent Cost*
  - *Inherent Contribution Preference*
  - *Explicit Reward*

- Model:
  - cost function $-cost_i^{UP}$
  - benefit function $gain_i^{UP}$

Contribution utility (u_i^{UP})

- Factors:
  - *Inherent Cost*
  - *Inherent Contribution Preference*
  - *Explicit Reward*
  - *Reputation / Feedback*
- Model:
  - cost function $-cost_i^{UP}$
  - benefit function $gain_i^{UP}$
  - sum of utility provided to other agents: $\sum_{j\neq i}^{n} v_j(a_i)$

Contribution utility (u_i^{UP})

- Factors:
  - *Inherent Cost*
  - *Inherent Contribution Preference*
  - *Explicit Reward*
  - *Reputation / Feedback*
- Model:
  - cost function $-cost_i^{UP}$
  - benefit function $gain_i^{UP}$
  - sum of utility provided to other agents: $\sum_{j \neq i}^{n} v_j(a_i)$
  - $\mathbb{E}(u_i^{UP}) = gain_i^{UP}(C) - cost_i^{UP}(C) + \sum_{j \neq i}^{n} v_j(a_i)$

Usenet utility model

Usenet utility model

- Assumptions:

- Formulæ:
  - Consumption:
    $u_i^{DN} = -cost_i^{DN}(C) + \sum_{j=1}^{m} w_{ij}\theta_i(size(C_j))$
  - Contribution:
    $u_i^{UP} = gain_i^{UP}(C) - cost_i^{UP}(C) + \sum_{j \neq i}^{n} v_j(a_i)$

Usenet utility model

- Assumptions:
  - linearity of cost and inherent gain
- Formulæ:
  - Consumption:
    $u_i^{DN} = \sum_{j=1}^{m}(w_{ij}\theta_i(size(C_j)) - \gamma_i size(C_j))$
  - Contribution:
    $u_i^{UP} = (\lambda_i - \gamma_i)size(C) + \sum_{j \neq i}^{n} v_j(a_i)$

Usenet utility model

- Assumptions:
  - linearity of cost and inherent gain
  - symmetry of θ_i

- Formulæ:
  - Consumption:
    $u_i^{DN} = \sum_{j=1}^{m}(w_{ij}\theta(size(C_j)) - \gamma_i size(C_j))$
  - Contribution:
    $u_i^{UP} = (\lambda_i - \gamma_i)size(C) + \sum_{j \neq i}^{n} v_j(a_i)$

Usenet utility model

- Assumptions:
  - linearity of cost and inherent gain
  - symmetry of θ_i
  - per-user partitioning of $\mathcal{Q}$

- Formulæ:
  - Consumption:
    $u_i^{DN} = \sum_{j=1}^{n}(b_{ij}\theta(c_j) - \gamma_i c_j)$
  - Contribution:
    $u_i^{UP} = (\lambda_i - \gamma_i)c_i + \sum_{j \neq i}^{n} v_j(a_i)$

Usenet utility model

- Assumptions:
  - linearity of cost and inherent gain
  - symmetry of θ_i
  - per-user partitioning of $\mathcal{Q}$
  - feedback is simply utility derived from the agent
- Formulæ:
  - Consumption:
    $u_i^{DN} = \sum_{j=1}^{n} \hat{v}_{ij}^{DN}, \quad \hat{v}_{ij}^{DN} = (b_{ij}\theta(c_j) - \gamma_i c_j)$
  - Contribution:
    $u_i^{UP} = (\lambda_i - \gamma_i)c_i + \sum_{j \neq i}^{n} \hat{v}_{ij}^{DN}$

Reputation un-motivated agents

Reputation un-motivated agents

- Assume that a_i has a bound on contribution k_i

Reputation un-motivated agents

- Assume that a_i has a bound on contribution k_i
- Then

$$\forall i,\, c_i = \begin{cases} k_i & \text{if } \lambda_i > \gamma_i \\ 0 & \text{otherwise} \end{cases}$$

is an equilibrium.

Reputation motivated agents

Reputation motivated agents

- For reputation-motivated agents:
  - There exist fixed c_i^* such that $\forall i$, $c_i = \min\{c_i^*, k_i\}$ is a unique Nash equilibrium.
  - Given θ there exists a threshold τ such that if

$$\sum_{j \neq i}^{n} b_{ij} \geq \tau \sum_{k}^{n} \gamma_k \tag{2}$$

then $c_i^* > 0$. Otherwise, $c_i^* = 0$.

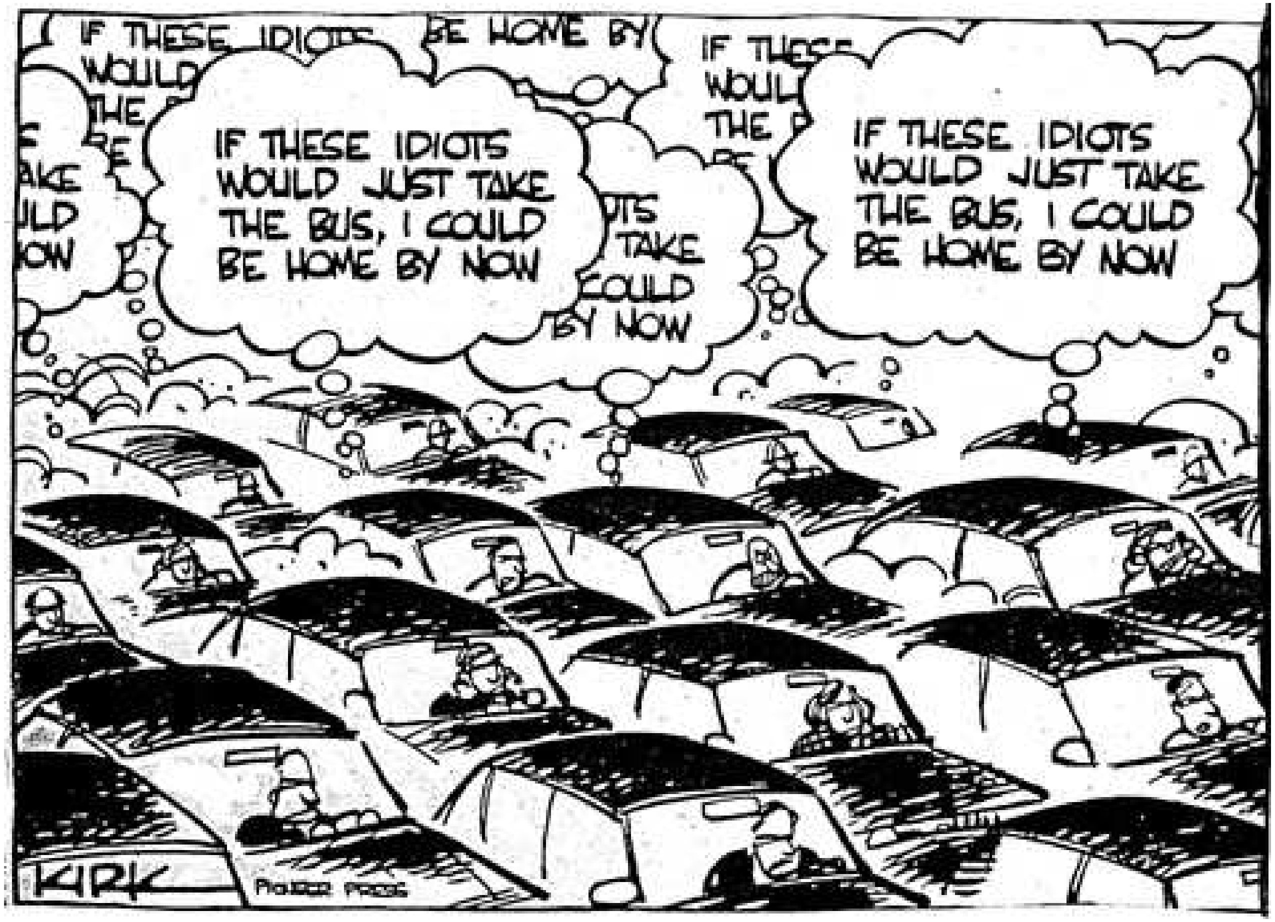
Reputation motivated agents

- For reputation-motivated agents:
  - There exist fixed c_i^* such that $\forall i$, $c_i = \min\{c_i^*, k_i\}$ is a unique Nash equilibrium.
  - Given θ there exists a threshold τ such that if

$$\sum_{j \neq i}^{n} b_{ij} \geq \tau \sum_{k}^{n} \gamma_k \tag{3}$$

    then $c_i^* > 0$. Otherwise, $c_i^* = 0$.
- Feedback in a group of users can regulate individual action to maximize collective welfare.

Congestion


IF THESE IDIOTS WOULD JUST TAKE THE BUS, I COULD BE HOME BY NOW
IF THESE IDIOTS WOULD JUST TAKE THE BUS, I COULD BE HOME BY NOW
KIRK
PIONEER PRESS

Resource competition game

Resource competition game

- Benefit matrix B has structure

$$\mathrm{B} = \begin{pmatrix} \blacksquare & 0 & \cdots & 0 \\ 0 & \blacksquare & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \blacksquare \end{pmatrix}$$

Resource competition game

- Benefit matrix B has structure

$$\mathrm{B} = \begin{pmatrix} \blacksquare & 0 & \cdots & 0 \\ 0 & \blacksquare & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \blacksquare \end{pmatrix}$$

- Collective action breaks down over entire system

Resource competition game

- Benefit matrix B has structure

$$\mathrm{B} = \begin{pmatrix} \blacksquare & 0 & \cdots & 0 \\ 0 & \blacksquare & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \blacksquare \end{pmatrix}$$

- Collective action breaks down over entire system
- Assume we have some finite resource with limit κ:
  - if sum of content less than κ, no change
  - otherwise, drop content until sum is less than κ

Resource competition (cont)

Resource competition (cont)

- Consider *groups* as players
  - action: upload $c_i \in [0, k_i]$ of content
  - utility: proportional to *non-dropped* content

Resource competition (cont)

- Consider *groups* as players
  - action: upload $c_i \in [0, k_i]$ of content
  - utility: proportional to *non-dropped* content
- Assume content is dropped uniformly:

$$u_i \propto \frac{\kappa c_i}{\sum_j c_j}$$

Resource competition (cont)

- Consider *groups* as players
  - action: upload $c_i \in [0, k_i]$ of content
  - utility: proportional to *non-dropped* content
- Assume content is dropped uniformly:

$$u_i \propto \frac{\kappa c_i}{\sum_j c_j}$$

- $\{k_1, k_2, \ldots, k_n\}$ is a Bayes-Nash equilibrium
  ... regardless of κ

Resource competition (cont)

- Consider *groups* as players
  - action: upload $c_i \in [0, k_i]$ of content
  - utility: proportional to *non-dropped* content
- Assume content is dropped uniformly:

$$u_i \propto \frac{\kappa c_i}{\sum_j c_j}$$

- $\{k_1, k_2, \ldots, k_n\}$ is a Bayes-Nash equilibrium
  ... regardless of κ
- Significant problem currently for Usenet servers

Contribution valuation

Contribution valuation

- Difficult to quantify eg., how to measure w_{ij}?, partitioning $\{Q_i\}$?

Contribution valuation

- Difficult to quantify eg., how to measure w_{ij}?, partitioning $\{Q_i\}$?
- What *can* be measured?
  - contributions: f_{up}
  - consumptions: f_{down}
  - size: f_{size}

Contribution valuation

- Difficult to quantify eg., how to measure w_{ij}?, partitioning $\{Q_i\}$?
- What *can* be measured?
  - contributions: f_{up}
  - consumptions: f_{down}
  - size: f_{size}
- All contributions of f are in the same class, so
  - $v(f) = \int \frac{f_{down}}{f_{up}} \cdot \theta(f_{size}) dt$

Global resource allocation

Global resource allocation

- Usenet: Bandwidth, retention

Global resource allocation

- Usenet: Bandwidth, retention
- Utility model for retention: sub-linear

Global resource allocation

- Usenet: Bandwidth, retention
- Utility model for retention: sub-linear
- Micro-economic vs. differential service

Global resource allocation

- Usenet: Bandwidth, retention
- Utility model for retention: sub-linear
- Micro-economic vs. differential service
- Current: manually per-group; volume-based

Global resource allocation

- Usenet: Bandwidth, retention
- Utility model for retention: sub-linear
- Micro-economic vs. differential service
- Current: manually per-group; volume-based
- Proposed: value over size

$$v(f)/f_{size}$$

Conclusions

Conclusions

- We have found that:

- Future directions:

Conclusions

- We have found that:
  - reputation can mitigate the effects of social dilemmas
- Future directions:

Conclusions

- We have found that:
  - reputation can mitigate the effects of social dilemmas
  - reputation is inadequate globally
- Future directions:

Conclusions

- We have found that:
  - reputation can mitigate the effects of social dilemmas
  - reputation is inadequate globally
- Explicit methods based on item value
- Future directions:

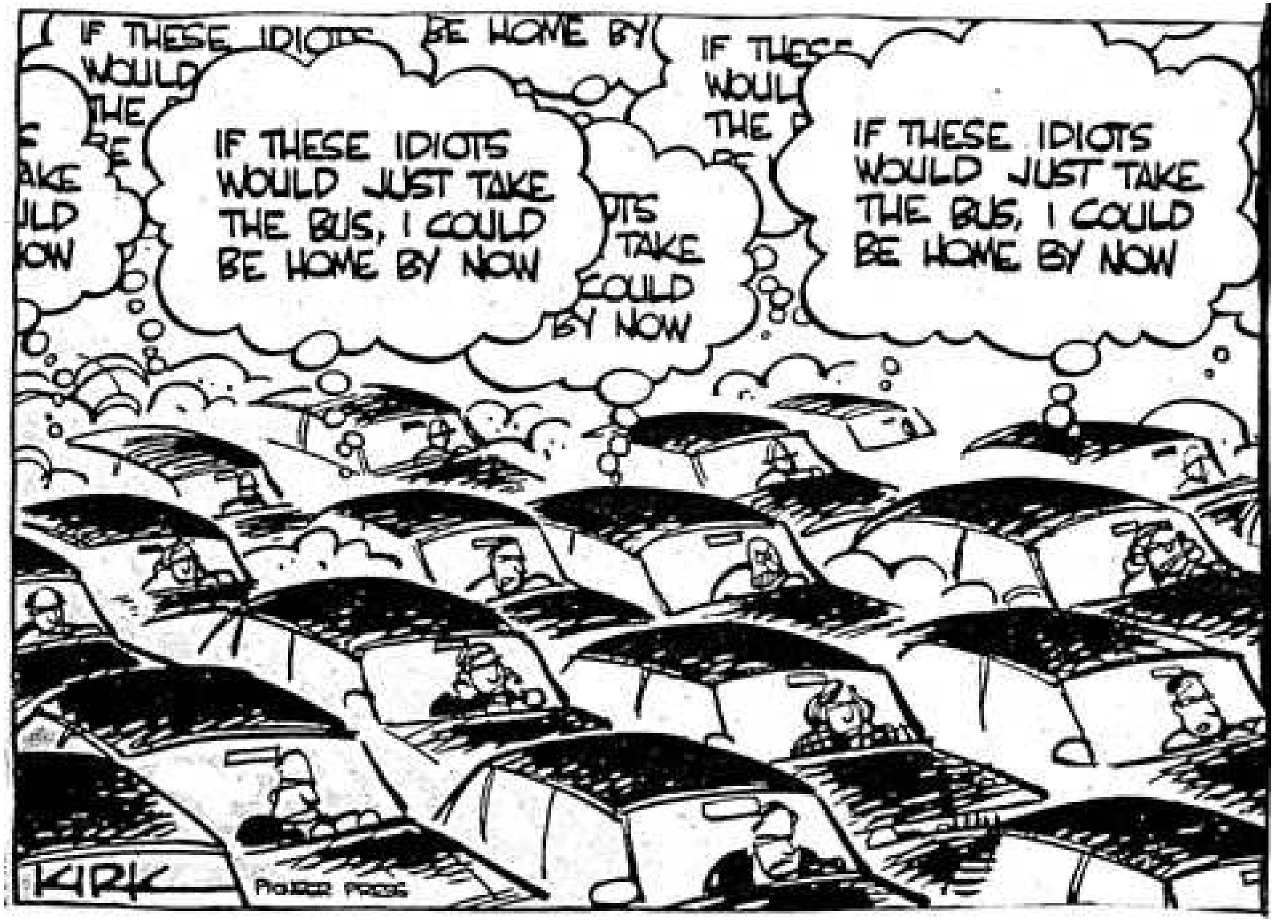
Conclusions

- We have found that:
  - reputation can mitigate the effects of social dilemmas
  - reputation is inadequate globally
- Explicit methods based on item value
- Future directions:
  - analysis of mix of reputation-sensitivity

Conclusions

- We have found that:
  - reputation can mitigate the effects of social dilemmas
  - reputation is inadequate globally
- Explicit methods based on item value
- Future directions:
  - analysis of mix of reputation-sensitivity
  - non-stationary repeated setting

Questions?


IF THESE IDIOTS WOULD JUST TAKE THE BUS, I COULD BE HOME BY NOW
IF THESE IDIOTS WOULD JUST TAKE THE BUS, I COULD BE HOME BY NOW
KIRK