

Invisible Sets: Mathematical Framework

kofi

1 Math Primer!

Hey everyone! Before we get into the paper's specifics, let's quickly make sure we're all on the same page with some fundamental math ideas. Economics has a ton of scary-looking symbols and syntax, but it's wayyy simpler than it seems :)

1. Sets: Collections of Stuff

- You already know sets from pre-calc! A set is just a collection of distinct objects.
- $i \in \{1, 2, \dots, N\}$: This means " i is one of the N individuals in our population."
- $C_i \subseteq \mathcal{X}$: This means " C_i is a subset of \mathcal{X} ." If \mathcal{X} is all possible choices (like every food item in the world), then C_i is a smaller collection (like the menu at a specific restaurant). So, an individual's **choice set** is just a smaller list of things they can actually pick from.

language

a 'subset' is a subordinate (i.e., "under") set; that is, a set within a set.

- 2^N : This notation means "the power set of N ." It's the set of **all possible subsets** of N . For this paper, it means "all possible social networks." Don't worry about it too much beyond that :)

2. Probability: Chance of Something Happening

- We write $\mathbb{P}(\text{event})$ to mean "the probability of some event happening." It's always a number between 0 and 1 (or 0% and 100%).
- Example: " $\mathbb{P}(\text{rain}) = 0.3$ " means there's a 30% chance of rain.
- $\mathbb{P}((i, j) \in \text{Network})$: This means "the probability that individual i and individual j are connected in the social network."

mathematical symbols

some of the symbols you might see:

- \in : "is an element of" (e.g., $x \in X$ means x is in the set X)
- \subseteq : "is a subset of" (e.g., $A \subseteq B$ means every element in A is also in B)
- \cup : "union" (e.g., $A \cup B$ is all elements in either A or B)
- \cap : "intersection" (e.g., $A \cap B$ is all elements in both A and B)
- $|$: "such that" or "given that" in probability (e.g., $P(A|B)$ is probability of A given B)

- $\mathbb{P}(T \in C_i | Z_i, g_i)$: Now this one has a pipe ‘|’ in it! That means “conditional probability.” It’s “the probability that treatment option T is available to individual i , *given* their observable characteristics Z_i and group membership g_i .” Think of it as: “What’s the chance of this happening, *if we already know this other thing is true?*”

3. Expected Value (\mathbb{E}): The Average Outcome

- Think of \mathbb{E} as the “average” or “long-run average” of something.
- If you flip a fair coin, the probability of heads is 0.5. If you assign “1” to heads and “0” to tails, the expected value of a flip is $\mathbb{E}[\text{outcome}] = 0.5 \times 1 + 0.5 \times 0 = 0.5$. It’s the weighted average of all possible outcomes, where the weights are their probabilities.
- $\mathbb{E}[y_i(1) - y_i(0)]$: This is really important in the paper! It means “the average difference in outcome if an individual *could* choose treatment 1 versus *could* choose treatment 0.” We’ll get deeper into “potential outcomes” later, but for now, just think average difference!
- $\mathbb{E}[y_i | x_i = 1, Z_i = z, g_i = g]$: This is the “average outcome y_i ”, *given that* individual i chose $x_i = 1$, *and given that* their characteristics are $Z_i = z$ and $g_i = g$.”

4. Argmax: Finding the Best

- $x_i = \arg \max_{x \in C_i} u_i(x, Z_i, \eta_i)$: This looks scary, but it’s just telling us *how* someone makes a choice. It means: “ x_i is the specific choice x (from the set of available choices C_i) that maximizes (gives the highest value to) the utility function u_i .”
- Think of it like picking the tastiest item on a menu. You look at all the available options (C_i), consider what you like (u_i), and pick the one that gives you the most happiness (maximizes utility).
- In this paper, we assume “deterministic choice with strict preferences,” which just means that for any given person, their utility function will always point to one unique, absolutely best choice from their available options. No ties or randomness in their decision making!

5. Summation (\sum)

- $\sum_{z,g}$: This is the capital Greek letter sigma¹, which means “sum.” If you see $\sum_{k=1}^N f(k)$, it means add up $f(1) + f(2) + \dots + f(N)$.
- $\sum_{z,g} \lambda_{z,g} \mathbb{E}[\dots]$: This means we’re adding up the expected values for each combination of z (observable characteristics) and g (group membership), weighted by $\lambda_{z,g}$ (which is the proportion of people in the population that have those specific z and g characteristics). It’s like calculating a total average by taking the average for each subgroup and then weighting them by how big each subgroup is.

6. Inequalities

- \geq (greater than or equal to) and \leq (less than or equal to). You know these!
- $\tau \in [\tau_L, \tau_U]$: This means “tau is in the interval from τ_L to τ_U .” So, tau is somewhere between τ_L and τ_U , inclusive.

¹...nvm

7. Functions

- $f : X \rightarrow Y$: Means “ f is a function that takes inputs from set X and produces outputs in set Y .”
- $u_i : \mathcal{X} \times \mathcal{Z} \times \mathcal{H} \rightarrow \mathbb{R}$: This means the utility function u_i takes three inputs (a choice x , observable characteristics Z_i , and unobserved heterogeneity η_i) and outputs a real number (\mathbb{R}), which is the utility value. This is aptly called the “utility function.”
- $f(x_i, Z_i, \varepsilon_i)$: A function that takes choices, observables, and random shocks to produce an outcome y_i .
- $\Gamma(Z_i, g_i, \mathcal{N}_i, \eta_i)$: The function that determines someone’s choice set based on their characteristics, group, network, and unobserved stuff.
- $h(\|Z_i - Z_j\|, |g_i - g_j|)$: A function h that takes distances in characteristics and group membership as inputs. The ‘ $\| \dots \|$ ’ and ‘ $| \dots |$ ’ are for “distance” or “absolute difference.”

8. Derivatives and Integrals (Briefly, for One Section!)

- $\frac{\partial}{\partial x} f(x, y)$: This is a “partial derivative.” If you have a function with multiple variables (like $f(x, y)$), a partial derivative tells you how f changes *only when x changes*, assuming y stays constant. It’s the slope of the function in one specific direction. Don’t worry too much about calculating these, just know what they mean!
- $\int f(x)dx$: This is an “integral.” It’s essentially finding the “area under the curve” of a function. It’s used to “sum up” values for continuous variables, like our z and g if they weren’t discrete categories. Again, don’t worry about calculating, just the concept.

Phew! You’ve got the basic math vocabulary now. Let’s get into the paper! :)

2 Model Setup

2.1 Basic Environment

Definition 1 (Choice Environment). *We consider a population of N individuals indexed by $i \in \{1, 2, \dots, N\}$. Each individual is characterized by:*

- Observable characteristics $Z_i \in \mathcal{Z}$ (what we can see, like age or income)
- Group membership $g_i \in \mathcal{G} = \{1, 2, \dots, G\}$ (like ‘student’ or ‘faculty’)
- Latent choice set $C_i \subseteq \mathcal{X}$ where \mathcal{X} is the universe of possible choices (the hidden list of options available to person i)
- Unobserved heterogeneity $\eta_i \in \mathcal{H}$ (hidden personal quirks, tastes, etc.)

Definition 2 (Choice Behavior). *Individual i chooses $x_i \in C_i$ to maximize utility:*

$$x_i = \arg \max_{x \in C_i} u_i(x, Z_i, \eta_i) \quad (1)$$

where $u_i : \mathcal{X} \times \mathcal{Z} \times \mathcal{H} \rightarrow \mathbb{R}$ is the utility function.

For simplicity, we assume deterministic choice with strict preferences, so the argmax is unique.

Definition 3 (Outcomes). *Individual i realizes outcome y_i according to:*

$$y_i = f(x_i, Z_i, \varepsilon_i) \quad (2)$$

where ε_i is an unobserved shock independent of all other variables.

key takeaway

People choose from a hidden set of options (C_i) to get the best outcome for themselves (u_i). We, the researchers, don’t get to see their C_i , u_i , η_i , or ε_i . This hidden info is the root of the problem! :)

2.2 Choice Set Formation

Assumption 1 (Choice Set Structure). *Choice sets are determined by:*

$$C_i = \Gamma(Z_i, g_i, \mathcal{N}_i, \eta_i) \quad (3)$$

where \mathcal{N}_i represents individual i 's social network and $\Gamma : \mathcal{Z} \times \mathcal{G} \times 2^N \times \mathcal{H} \rightarrow 2^{\mathcal{X}}$ is the choice set formation function.

Assumption 2 (Network Homophily). *The probability that individuals i and j are connected decreases in their distance:*

$$\mathbb{P}((i, j) \in \text{Network}) = h(\|Z_i - Z_j\|, |g_i - g_j|) \quad (4)$$

where h is decreasing in both arguments.

2.3 Researcher's Problem

The researcher observes data $\{(x_i, y_i, Z_i, g_i)\}_{i=1}^N$ but does not observe $\{C_i, \eta_i, \varepsilon_i\}_{i=1}^N$.

Definition 4 (Treatment Effect of Interest). *For a binary treatment $T \in \{0, 1\} \subset \mathcal{X}$, the average treatment effect is:*

$$\tau = \mathbb{E}[y_i(1) - y_i(0)] \quad (5)$$

where $y_i(t)$ is the potential outcome when individual i chooses treatment level t .

Definition 5 (Identification). *The treatment effect τ is identified if there exists a unique value τ^* such that for all data generating processes consistent with the observed data, $\tau = \tau^*$.*

understanding identification

Can we calculate τ exactly from the data we can see? If yes, it's identified. If no (because different "hidden truths" could lead to the same data), it's not! Think of it like trying to find the individual prices of apples and bananas if you only ever see the total bill. If there are multiple combinations of prices that give the same total, you can't uniquely identify the individual prices.

3 Main Theoretical Results

3.1 Non-Identification Result

Theorem 1 (Non-Identification of Treatment Effects). *When choice sets $\{C_i\}_{i=1}^N$ are unobserved and vary across individuals, the treatment effect τ is not identified, even with infinite data.*

Proof. We construct two data generating processes (DGPs) that produce identical observable data but have different treatment effects.

Construction: Consider two individuals with identical observables $(Z_1, g_1) = (Z_2, g_2) = (z, g)$ but different choice sets:

$$C_1 = \{0, 1\} \quad (\text{full choice set}) \quad (6)$$

$$C_2 = \{1\} \quad (\text{restricted choice set}) \quad (7)$$

Parallel: Two students with identical grades and high school (same Z, g). Student 1 can apply to any college (C_1). Student 2 can only apply to one specific local college (C_2).

DGP 1: Assume utility functions such that:

$$u_1(1, z, \eta_1) > u_1(0, z, \eta_1) \Rightarrow x_1 = 1 \quad (8)$$

$$u_2(1, z, \eta_2) > u_2(0, z, \eta_2) \Rightarrow x_2 = 1 \quad (9)$$

In this “true world,” both individuals chose treatment 1, and crucially, even if person 2 could have chosen 0, they would have preferred 1. Both individuals choose treatment, with potential outcomes:

$$y_1(0) = 0, \quad y_1(1) = 1 \quad \Rightarrow \tau_1 = 1 \quad (10)$$

$$y_2(0) = 0, \quad y_2(1) = 1 \quad \Rightarrow \tau_2 = 1 \quad (11)$$

Average treatment effect: $\tau^{(1)} = 1$.

DGP 2: Same choice behavior but different potential outcomes:

$$y_1(0) = 0, \quad y_1(1) = 1 \quad \Rightarrow \tau_1 = 1 \quad (12)$$

$$y_2(0) = 1, \quad y_2(1) = 1 \quad \Rightarrow \tau_2 = 0 \quad (13)$$

Here, person 2 chose 1 (because it was their only option), but if they could have chosen 0, their outcome would have been the same. So, for them, the treatment had no effect ($\tau_2 = 0$). Average treatment effect: $\tau^{(2)} = 0.5$.

Observable Equivalence: In both DGPs, we observe $(x_1, y_1) = (1, 1)$ and $(x_2, y_2) = (1, 1)$ with identical (Z_i, g_i) . The observed data distributions are identical. **Different Treatment Effects:** $\tau^{(1)} = 1 \neq 0.5 = \tau^{(2)}$.

Since two different treatment effects are consistent with the same observed data, τ is not identified. \square

why non-identification is a big deal

This theorem is the heart of the paper. It says that when people have different options that we can't see, it's fundamentally impossible to figure out the true average impact of a ‘treatment’. Why? Because we don't know if someone chose something because they preferred it (true effect) or if it was their only option (selection bias). The outcomes might look the same, but the reasons for the choices (and thus the true effect) are different!

3.2 Sharp Bounds Result

Theorem 2 (Sharp Bounds Under Minimal Assumptions). *Under Assumptions 1 and the following restrictions:*

(i) *Support condition:* $\mathbb{P}(T \in C_i | Z_i, g_i) \in [\underline{p}(Z_i, g_i), \bar{p}(Z_i, g_i)]$

(ii) *Monotonic treatment response:* $y_i(1) \geq y_i(0)$ for all i

The treatment effect is bounded by:

$$\tau \in [\tau_L, \tau_U] \quad (14)$$

where the bounds are given by:

$$\tau_L = \sum_{z,g} \lambda_{z,g} [\underline{p}(z, g) \cdot \mathbb{E}[y_i | x_i = 1, Z_i = z, g_i = g]] \quad (15)$$

$$\tau_U = \sum_{z,g} \lambda_{z,g} [\bar{p}(z, g) \cdot \mathbb{E}[y_i | x_i = 1, Z_i = z, g_i = g]] \quad (16)$$

and $\lambda_{z,g} = \mathbb{P}(Z_i = z, g_i = g)$. Moreover, these bounds are sharp.

Proof. Step 1: Decomposition of Treatment Effect The average treatment effect can be written as:

$$\tau = \mathbb{E}[y_i(1) - y_i(0)] \quad (17)$$

$$= \sum_{z,g} \lambda_{z,g} \mathbb{E}[y_i(1) - y_i(0) | Z_i = z, g_i = g] \quad (18)$$

Step 2: Conditional Treatment Effect Bounds For individuals with characteristics (z, g) , we can bound the conditional treatment effect. Under monotonicity, we assume for the lower bound that the

treatment effect is minimal (e.g., $y_i(0)$ is close to $y_i(1)$ for those who chose treatment) and access is at \underline{p} . For the upper bound, we assume the treatment effect is maximal (e.g., $y_i(0) = 0$) and access is at \bar{p} .

Step 3: Aggregation Summing over all (z, g) pairs weighted by population shares gives the stated bounds.

Step 4: Sharpness The bounds are achieved by assuming the most restrictive or most permissive choice set structures consistent with observations. \square

bounds: getting a grip on uncertainty

If we can't know the exact answer, knowing the range of possible answers is the next best thing! This theorem tells us that with two reasonable assumptions (some knowledge of treatment availability and that treatment doesn't make things worse), we can calculate a range for our average treatment effect τ .

3.3 Network Effects on Identification

Theorem 3 (Homophily Worsens Identification). *Under Assumptions 1-2, network homophily reduces the width of bounds $\tau_U - \tau_L$ achievable through variation in choice sets, thereby worsening identification.*

Proof. Setup: Consider a network where connection probability between i and j is:

$$p_{ij} = \exp(-\alpha\|Z_i - Z_j\| - \beta|g_i - g_j|) \quad (19)$$

Step 1: Choice Set Correlation As homophily increases ($\alpha, \beta \rightarrow \infty$), choice sets become perfectly correlated within groups:

$$\rho_C(z, g) \rightarrow 1 \quad (20)$$

Step 2: Variance Reduction The variance in choice set access within group (z, g) approaches zero as $\rho_C(z, g) \rightarrow 1$. This means choice sets become identical within groups (everyone has the option or no one does).

Step 3: Identification Impact When choice sets are identical within groups, it becomes impossible to tell if someone chose something because they *wanted* it (true treatment effect) or because it was their *only option* (selection effect). This leads to wider, less useful bounds for τ .

Step 4: Welfare Implications Homophily leads to "allocative inefficiency." Treatment isn't necessarily allocated to those who would benefit most, but rather to those connected in networks. This causes "welfare loss." \square

homophily is bad for identification

"Birds of a feather flock together" actually makes our research job more interesting! If social circles are too segregated, people within a group will have very similar access to opportunities. This makes it tough to know if an observed good outcome is due to the treatment itself, or just because only the "right" people (who would succeed anyway) got access. This means more uncertainty in our estimates and potentially means good opportunities don't reach those who need them most.

4 Extensions and Robustness

4.1 Dynamic Choice Sets

Definition 6 (Dynamic Choice Set Formation). *In period t , individual i 's choice set evolves according to:*

$$C_{it} = \Gamma(Z_{it}, g_i, \mathcal{N}_{it}, C_{i,t-1}, x_{i,t-1}, \eta_{it}) \quad (21)$$

where past choices $x_{i,t-1}$ and choice sets $C_{i,t-1}$ affect current opportunities.

Proposition 1 (Dynamic Identification). *In the dynamic setting, treatment effects are not identified even under stronger assumptions about choice set formation, due to path dependence in opportunity access.*

Proof Sketch. Past choices affecting current choice sets create an additional layer of unobservable differences, compounding the static identification problem. \square

4.2 Continuous Treatment Intensity

Definition 7 (Continuous Treatment). *Let $\mathcal{X} = [0, \bar{x}]$ represent continuous treatment intensity (e.g., dosage of medicine), and choice sets be intervals $C_i = [0, c_i]$ where c_i is the maximum treatment intensity available to individual i .*

Proposition 2 (Continuous Treatment Bounds). *Under continuous treatment with interval choice sets, the treatment effect bounds become:*

$$\tau_L = \int \underline{c}(z, g) \cdot \frac{\partial}{\partial x} \mathbb{E}[y_i | x_i = x, Z_i = z, g_i = g] \Big|_{x=0} dF(z, g) \quad (22)$$

$$\tau_U = \int \bar{c}(z, g) \cdot \frac{\partial}{\partial x} \mathbb{E}[y_i | x_i = x, Z_i = z, g_i = g] \Big|_{x=0} dF(z, g) \quad (23)$$

where $\underline{c}(z, g)$ and $\bar{c}(z, g)$ are bounds on maximum available treatment intensity.

4.3 Welfare Analysis

Definition 8 (Social Welfare). *Define social welfare as:*

$$W = \sum_{i=1}^N \mathbb{E}[y_i | C_i, Z_i, g_i] \quad (24)$$

Proposition 3 (Optimal Information Provision). *The social planner's problem is to choose information provision policy $\pi : \mathcal{Z} \times \mathcal{G} \rightarrow [0, 1]$ that maximizes expected welfare subject to budget constraint $\sum_{z,g} \lambda_{z,g} \pi(z, g) \leq B$. The optimal policy satisfies:*

$$\pi^*(z, g) = \begin{cases} 1 & \text{if } \frac{\partial W}{\partial \pi(z, g)} \geq \lambda \\ 0 & \text{if } \frac{\partial W}{\partial \pi(z, g)} < \lambda \end{cases} \quad (25)$$

where λ is the Lagrange multiplier on the budget constraint.

5 Testable Implications

5.1 Information Experiments

Proposition 4 (Information Intervention Effects). *If choice sets are constrained by information, then random information provision should increase treatment take-up. The magnitude of increase identifies the extent of information constraints. Formally, let $I_i \in \{0, 1\}$ be a random information treatment. Then:*

$$\mathbb{E}[x_i | I_i = 1, Z_i, g_i] - \mathbb{E}[x_i | I_i = 0, Z_i, g_i] = \Delta(Z_i, g_i) \quad (26)$$

where $\Delta(Z_i, g_i) > 0$ indicates information constraints for group (Z_i, g_i) .

5.2 Referral Pattern Tests

Proposition 5 (Network Clustering Test). *Under network-based choice set formation, treatment choices should exhibit excess clustering within demographic groups, conditional on observables. This can be tested using the statistic:*

$$T_{cluster} = \frac{1}{N} \sum_{i=1}^N \sum_{j \in \mathcal{N}_i} |x_i - x_j| - \mathbb{E}[|x_i - x_j| | Z_i, Z_j, g_i, g_j] \quad (27)$$

where negative values indicate excess clustering consistent with network effects.

5.3 Bounds Tightening

Proposition 6 (Information Value). *As researchers collect better data on choice set formation (through surveys, experiments, etc.), the bounds should tighten. This can be formalized as:*

$$\tau_U(\mathcal{I}') - \tau_L(\mathcal{I}') \leq \tau_U(\mathcal{I}) - \tau_L(\mathcal{I}) \quad (28)$$

where $\mathcal{I} \subset \mathcal{I}'$ represents information sets about choice set formation.

6 Computational Procedures

6.1 Bounds Estimation Algorithm

1. **Data Preparation:** Group your data by observable characteristics (Z_i, g_i) .
2. **Within-Group Analysis:** For each group:
 - (a) Estimate the average outcome for those who chose the treatment $(\mathbb{E}[y_i | x_i = 1, \dots])$.
 - (b) Specify the lowest and highest likely probabilities of treatment access (\underline{p}, \bar{p}) for that group.
3. **Bounds Computation:** Plug these numbers into the formulas from Theorem 2 to get τ_L and τ_U .
4. **Confidence Intervals:** Use statistical methods (like bootstrapping) to estimate the uncertainty around your calculated bounds.

6.2 Network Effects Estimation

For network-based choice set formation:

1. **Network Construction:** Build a map of who is connected to whom.
2. **Homophily Measurement:** Estimate how strongly people connect with similar others.
3. **Choice Set Correlation:** See if people within the same groups (especially if connected) make similar choices.
4. **Welfare Analysis:** Run computer simulations to see how different network structures might change overall welfare.

7 Conclusion

This mathematical framework provides a solid way to analyze problems when people's options are hidden. The key takeaways are:

1. **Non-identification result:** You can't pinpoint exact treatment effects when choice sets are hidden and vary.
2. **Sharp bounds:** But you can find a useful range for the effect under certain assumptions.
3. **Network effects:** Social connections can make it harder to identify effects and lead to worse overall outcomes.
4. **Policy implications:** This framework helps design interventions, like providing better information or structuring platforms.

The framework is flexible and can be applied to many areas (education, jobs, tech adoption) while also giving practical steps for researchers.

ready for action!

You’ve made it! You’ve walked through the entire paper, from basic definitions to deep theoretical results and practical applications. You’re now armed with the mathematical understanding and terminology to tackle the core of this policy paper’s mathematical section.

Now, for some exercises to solidify your understanding. Don’t worry, they’re designed to help you think through the concepts, not stump you! :)

Exercises to Get You Thinking! :)

Exercise 1: Decoding the Language

1. Explain in your own words what “latent choice set C_i ” means. Give a real-world example not used in this guide. Why is it a problem that it’s “latent”?
2. What does $\mathbb{E}[y_i(1) - y_i(0)]$ mean? Why is it called a “treatment effect”?
3. If a researcher states that a certain effect is “not identified,” what does that practically mean for their ability to draw conclusions?

Exercise 2: Probability and Expected Value Warm-up

Let’s say we have a tiny population of 3 individuals.

- Individual 1: $Z_1 = \text{young}, g_1 = \text{student}$
- Individual 2: $Z_2 = \text{old}, g_2 = \text{student}$
- Individual 3: $Z_3 = \text{young}, g_3 = \text{faculty}$

The treatment T is ‘attending a workshop’. Outcomes y_i are 1 (got a promotion) or 0 (didn’t).

1. What are the values of $\lambda_{\text{young}, \text{student}}$, $\lambda_{\text{old}, \text{student}}$, and $\lambda_{\text{young}, \text{faculty}}$? (Hint: these are just population proportions!)
2. Suppose you observe the following:
 - Individual 1 attended the workshop ($x_1 = 1$) and got a promotion ($y_1 = 1$).
 - Individual 2 attended the workshop ($x_2 = 1$) and got a promotion ($y_2 = 1$).
 - Individual 3 did NOT attend the workshop ($x_3 = 0$) and did NOT get a promotion ($y_3 = 0$).

Calculate $\mathbb{E}[y_i | x_i = 1, Z_i = \text{young}, g_i = \text{student}]$ based on this tiny sample.

3. If you somehow knew that for all ‘young students’, $\mathbb{P}(T \in C_i | Z_i = \text{young}, g_i = \text{student})$ was between 0.8 and 1.0 (meaning they almost always had the option), and for ‘old students’ it was between 0.1 and 0.2, and for ‘young faculty’ it was between 0.0 and 0.05, how would this information, combined with the assumption of monotonic treatment response, affect your ability to estimate the overall treatment effect? (No need to calculate exact bounds, just describe the intuition!)

Exercise 3: Non-Identification in Action

Imagine you're studying the effect of owning a smartphone ($T = 1$) versus not ($T = 0$) on people's daily happiness (y_i). You observe two people:

- Person A: owns a smartphone, very happy.
- Person B: owns a smartphone, very happy.

You don't know their choice sets.

1. Can you construct two different "true worlds" (DGPs) that lead to these same observed outcomes but imply different average effects of smartphone ownership on happiness? (Hint: Think about what their hidden choice sets C_i and potential outcomes $y_i(0), y_i(1)$ might be in each world).
2. Why does the non-identification theorem hold even with "infinite data" in this smartphone example?

Exercise 4: Homophily's Impact

You're studying a peer mentoring program in high schools.

- Students are more likely to be friends with (and thus influenced by) other students in the same grade and with similar academic performance.
 - If a student's friends join the mentoring program, the student is more likely to join, and more likely to have access to the program.
1. How does this situation demonstrate "network homophily"?
 2. Based on Theorem 3, how would this homophily affect your ability to identify the true effectiveness of the mentoring program on student grades? Explain intuitively.
 3. How might this homophily lead to "welfare loss" in the context of the mentoring program?

Exercise 5: Policy Implications

Based on the paper's insights:

1. If a local government wants to increase participation in a recycling program, but suspects people don't know about collection points, what kind of "information experiment" (as described in section 4.1) could they run? What would a successful outcome look like?
2. Imagine a new online platform designed to connect job seekers with employers. How could the principles of network effects (section 2.3) and optimal information provision (section 3.3) guide the design of this platform to maximize social welfare (i.e., get more people into suitable jobs)?

-kofi