A Nudge-Driven Model of Decision Effort and System Choice

Medhavi Devraj and Siddharth Shrimal

April 29, 2025

Ashoka University

Today's Path

Introduction

Research Question

Context

- Decision-making often involves balancing mental effort against the risk of making a mistake.
- Nudges can influence this balance.
- Studying how nudges affect thinking styles is important for understanding when individuals choose to rely on intuition versus deliberate reasoning in the presence of nudges.

Formal Research Question

How do nudges interact with cognitive effort and decision accuracy to influence different cognitive environments?

Literature Anchors & Model Links

Kahneman- System 1 and System 2 Thinking

Take-aways

- Kahneman made a distinction between two types of decision making systems: System 1 and System 2
- System 1 runs automatically—fast, effortless, emotional; System 2 is the manual gear—slow, deliberate, and resource-hungry.
- Relying on System 1 saves mental energy but invites biases; invoking System 2 improves accuracy but costs time and effort.

Model lens

Endogenous Accuracy

There is ε that measures the effort made in a decision which impacts $\mathbb{P}(\textit{of being wrong})$

Diminishing returns explain why most choices stay intuitive.

Nudge Toolkit $\rightarrow$ Tweaking θ and ε

Thaler and Sunstein had defined nudges in the most apt way: *A nudge... is any aspect of the choice architecture that alters people's behavior in a predictable way without... significantly changing their economic incentives.*

- A choice decision arises out of expected utility minus the effort expended.
- We incorporate nudges in such a manner that neither the effort expended nor the payoff from choice is impacted

Model Lens

Nudge Parameter ψ

$$f(\varepsilon, \psi) \quad r(\psi)$$

Madrian & Shea – Automatic 401(k) Enrolment

Take-aways

- Natural experiment: firm switched *only* the enrolment rule from *opt-in* to *opt-out*; match, fees and fund menu stayed constant.
- New-hire participation rate leapt **37 %** $\rightarrow$ **86 %**, while long-tenured employees remained flat at $\approx 49\,\%$.
- Shows a default nudge transforms low-effort ("System 1") decisions but leaves high-effort ("System 2") decisions almost untouched.

Model lens

Asymmetric -effect

We encode the default as a nudge ψ such that

$$f(\varepsilon', \psi) = f(\varepsilon') + \Delta f_{S1} \quad \text{(System 1 accuracy rises)}$$

$$f(\varepsilon) \text{ has } \textbf{no } \psi \text{ term} \quad \text{(System 2 accuracy unchanged)}$$

and System 2 bears a lower effective cost $\varepsilon - r(\psi)$.

Formal Framework

Decision Environment

- Say an individual has to choose between two choices x and y that yield utilities $U(x)$ and $U(y)$ respectively.
- Such a decision can be taken with System 1 thinking or System 2 thinking
- The right choice is denoted as $max(U(x), U(y))$ and the wrong choice is denoted as $min(U(x), U(y))$

Probability of Getting the Decision Right

Depending on the type of system used for making a decision- system 1 or system 2, there is some amount of effort expended

Effort and Probability

The model is created in such a way that the $\mathbb{P}(of\ being\ right)$ in endogenous and dependent on effort expended

$$\varepsilon' = \text{effort expended in system 1 thinking}$$

$$\varepsilon = \text{effort expended in system 2 thinking}$$

$$\text{Trivially, } \varepsilon > \varepsilon'$$

$$f(\varepsilon) = \mathbb{P}(of\ being\ right) \text{ in system 2}$$

$$f(\varepsilon') = \mathbb{P}(of\ being\ right) \text{ in system 1}$$

The behaviour of f is such that $f'_\varepsilon > 0$ and $f''_\varepsilon < 0$ hence it's an increasing and convex function on ε

Psi and the Introduction of Nudge

Like we explained earlier, nudges do not impact the utility of outcomes. Hence, we introduce a nudge variable ψ into the model. ψ impacts the model in a twofold manner:

System	**Probability**	**Effort**
System 1 (low effort)	$f(\varepsilon', \psi)$ — accuracy changes with ψ	ε' — effort remains fixed; ψ does not raise work.
System 2 (high effort)	$f(\varepsilon)$ — no direct effect of ψ	$\varepsilon - r(\psi)$ — ψ effects part of cognitive effort without changing accuracy.

Step 1: Activate a system

Now, we want to understand situations in which people use system 2 over system 1, or vice versa. For that, let us lay down the utility from using system 1 and system 2

Utilities without Nudges

$$\mu_{S1} = f(\varepsilon')\ \max(U(x), U(y)) + [1 - f(\varepsilon')]\ \min(U(x), U(y)) - \varepsilon'$$

$$\mu_{S2} = f(\varepsilon)\ \max(U(x), U(y)) + [1 - f(\varepsilon)]\ \min(U(x), U(y)) - \varepsilon$$

System 2 is used when:

$$\mu_{S2} > \mu_{S1} \quad \Rightarrow \quad \varepsilon - \varepsilon' < |U(x) - U(y)|[f(\varepsilon) - f(\varepsilon')]$$

Intuitively

$$\textit{Difference in effort} <$$
$$(\textit{Cost of being wrong}) \times (\textit{Difference in } \mathbb{P}(\textit{of being wrong}))$$

Step 2: Effect of a nudge (in the activated system)

Now, introduce the nudge variable ψ into the model.

Utilities with Nudges

$$\mu_{S1} = f(\varepsilon', \psi)\ \max(U(x), U(y)) + [1 - f(\varepsilon', \psi)]\ \min(U(x), U(y)) - \varepsilon'$$

$$\mu_{S2} = f(\varepsilon)\ \max(U(x), U(y)) + [1 - f(\varepsilon)]\ \min(U(x), U(y)) - (\varepsilon - r(\psi))$$

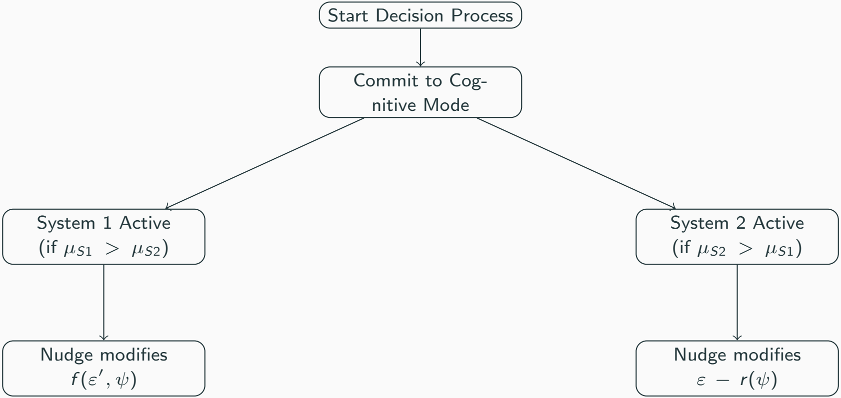
If one is using system 1 thinking apriori, and a nudge is introduced, the new condition becomes:

$$\varepsilon - \varepsilon' > |U(x) - U(y)|[f(\varepsilon) - f(\varepsilon')] \rightarrow \varepsilon - \varepsilon' > |U(x) - U(y)|[f(\varepsilon - f(\varepsilon', \psi)]$$

Similarly, if one is using system 2 thinking apriori, and a nudge is introduced, the new condition becomes:

$$\varepsilon - \varepsilon' - r(\psi) < |U(x) - U(y)|[f(\varepsilon) - f(\varepsilon')] \rightarrow \varepsilon - \varepsilon' > |U(x) - U(y)|[f(\varepsilon - f(\varepsilon')]$$

Sequence



Start Decision Process
Commit to Cog-
nitive Mode
System 1 Active
(if $\mu_{S1} > \mu_{S2}$)
System 2 Active
(if $\mu_{S2} > \mu_{S1}$)
Nudge modifies
$f(\varepsilon', \psi)$
Nudge modifies
$\varepsilon - r(\psi)$

Cases and Outcomes

Cases

There can be several cases for the nudge to work:

1. You're using system 1 thinking. Hence $\mu_{S1} > \mu_{S2}$
   1.1 Nudge works in the "right" direction i.e; nudging you towards the correct choice
   1.2 Nudge works in the "wrong" direction i.e; nudging you towards the incorrect choice
2. You're using system 2 thinking. Hence $\mu_{S1} < \mu_{S2}$
   2.1 Nudge works in the "right" direction i.e; nudging you towards the correct choice
   2.2 Nudge works in the "wrong" direction i.e; nudging you towards the incorrect choice

How does the nudge work- Case 1 — System 1 thinking

We have then $\mu_{S1} > \mu_{S2}$ gives

$$\varepsilon - \varepsilon' \ > \ |u(x) - u(y)| \left[f(\varepsilon) - f(\varepsilon', \psi)\right]. \quad (1)$$

Importantly, note that $r(\psi)$ does not show up here, because the nudge has no effect on system 2's effort. If the decision making is apriori, system 1, we will not include the effect of the nudge on system 2' effort

1. **If the nudge works in the right direction** $(\psi \uparrow)$, then and $f(\varepsilon'; \psi) \uparrow$. Consequently, the right-hand side (RHS) decreases; the agent stays in system 1 but with a higher probability of choosing correctly.
2. **If the nudge works in the wrong direction** $f(\varepsilon'; \psi) \downarrow)$, the RHS increases.

How does the nudge work- Case 1 — System 1 thinking

Possibility	**Condition**	**Intuition**
(a) You still stay in S1	New RHS is *still* $<$ LHS	The nudge isn't strong enough to push you into attentive (System 2) thinking.
(b) You switch to S2	New RHS $>$ LHS	The cue now looks wrong enough that the added accuracy is worth paying S2's cognitive cost, so you switch and remain attentive.

How does the nudge work- Case 2 — System 2 thinking

We have then $\mu_{S1} < \mu_{S2}$ gives

$$\varepsilon - \varepsilon' - r(\psi) \; < \; |u(x) - u(y)| \left[f(\varepsilon) - f(\varepsilon') \right]. \tag{2}$$

Here, like in the previous comment, if the decision is apriori in system 2, the nudge will not impact the probability of system 1, hence we have $f(\varepsilon')$ and not $f(\varepsilon', \psi)$

If the nudge works in the right direction Then $\psi \uparrow$, and the LHS $\downarrow$. Hence, you're even more inclined to stay in system 2

If the nudge works in the wrong direction Then $\psi \downarrow$, and the LHS $\uparrow$.

Case 2

Possibility	Condition	Intuition
(a) You still stay in S2	New LHS is *still* $<$ fixed RHS	The smaller refund is annoying, but not enough to offset System 2's higher expected-accuracy benefit.
(b) You drop to S1	New LHS $>$ RHS	The extra effort of System 2 is no longer justified, so it is rational to switch down to System 1.

In short

System 1 (low effort, fast thinking)

- **Helpful nudges** raise the accuracy of System 1, making it more attractive to stay intuitive.
- **Harmful nudges** reduce System 1 accuracy; if the perceived risk of mistake grows large enough, individuals rationally switch to System 2.

System 2 (high effort, deliberate thinking)

- **Helpful nudges** lower the effective cognitive cost $(\epsilon - r(\psi))$, encouraging individuals to stay in System 2.
- **Harmful nudges** (smaller $r(\psi)$) increase perceived cognitive cost; if the accuracy benefit no longer justifies the effort, individuals may switch back to System 1.

An Interesting Dynamic Within System 1

Threshold for switching out of System 1 under a harmful nudge:

For a bad cue, $\Delta f_{S1}(\psi) < 0$. Taking absolute values, the critical threshold is:

$$\boxed{|\Delta f_{S1}|_{\text{crit}} = \frac{\epsilon - \epsilon'}{|u(x) - u(y)|} - [f(\epsilon) - f(\epsilon')]}$$

Intuition:

- **High stakes** ($|u(x) - u(y)|$ large) $\Rightarrow$ denominator big $\Rightarrow$ threshold small.
- **Low Stakes** ($|u(x) - u(y)|$ small) $\Rightarrow$ denominator small $\Rightarrow$ threshold large.

So a harmful nudge works better in a low stakes environment?

Conclusion

Our model provides a structured answer to the core question: *"How do nudges interact with cognitive effort and decision accuracy to influence different cognitive environments?"*

Understanding how nudges interact with cognitive effort — across System 1 and System 2 thinking — is crucial for designing interventions that genuinely improve decision-making.

References

References

- Madrian, B.C. and Shea, D.F. (2001). *The Power of Suggestion: Inertia in 401(k) Participation*. Quarterly Journal of Economics.
- Kahneman, D. (2003). “Maps of Bounded Rationality.” *American Economic Review*, 93(5).
- Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.