Value, Choice, and Experimental Trolleys

Frank P. DeVita

January 2026

Value plays an important role in perceptual judgement and decision. Even the simplest choices entail, reflect, or are influenced by values. The coffee one drinks reflects their preferred taste, and the clothes they wear reflect a valued style. One's beliefs and desires reflect what one values. More complicated decisions are also value laden. The endeavors one pursues express what they consider important, and moral values are reflected in one's actions toward others.

Value *sets the context* for decisions by centering one's judgements and actions to a standard[1]. Without value, decisions would be listless. There would be little reason or motivation to select one option among many, nor many useful guides for decision making. By providing a background on which choices are made, value determines *why* an agent chooses the way they do, and so *which* choices they make. In this paper, I explore the nature of, and relation between, value and choice from philosophical and

[1] Unger 1996.

neural standpoints, suggest an experiment that aims to discriminate the role of value in complex decisions, and consider interpretations of its possible results.

Value and Decision

Value is an often discussed yet nebulously defined concept in the sciences and philosophy[2]. We value things like money, stability, heirlooms, flags, other people, health, and our independence. Value is a kind of *bias*—it drives deciders to prefer some outcomes or things over others and to make different choices than they would otherwise. Value therefore plays a determining role in decision making. If I value my health, I won't choose processed foods, while someone who values the lowest food cost might. I might find cigarettes unhealthy and distasteful and so not smoke them, yet someone else might find them relaxing and pleasurable and develop a smoking habit. *Value* determines these differences in decision and action. Decision thus not only depends on the world and the options it presents, but also on facts about the agent making the decision.

Factors that determine a decision can be *external* features of a stimulus or the environment, or *internal* features of the decider. External factors might be the quality of a stimulus or a quantity of available options, while internal factors might be the amount

[2] Glimcher 2014, Hume 1739, Levy 2012, Nguyen2025, Glimcher 2023.

of attention given to a stimulus or the subset of options a decider considers relevant. Value shapes and influences the latter factors, and shapes decision making *from the inside* in a manner analogous to how a noisy stimulus or environmental change might shape decision *from the outside*. Sets of options exist *out there* but their presentation to a decision agent is colored by value and other internal factors, and so is choice behavior.

The set of factors a decision making study focuses on matters. Focusing on external factors and perceptual judgement will yield data and conclusions about the ability of subjects to *report* their sensations and perceptual states—their mental states—accurately. The decisions being observed and measured are about links between mental states and the world, namely the congruence between visual experiences and visual stimuli. The choice a subject makes in an external factor-focused experiment is thus about the relation between the perceptual system and the external world *via* the decider's ability to report on their own mental states. Focusing on internal factors and value will yield different data and conclusions about the deliberative process, namely the decider's idiosyncratic preferences and their fluctuations as a function of stimuli, often point or monetary values.

Focusing on either internal or external factors in decision making research results in an incomplete picture of the decision making process that misses aspects of *what it's like* to make a decision. The *experience* of decision making entails perceiving the world, cognizing the available options, weighing them, and deciding to act. The perceptual

judgement approach concentrates on the bookends of the decision experience by relying on stimulus presentation and reporting via action. The value-based approach captures the middle and end of decision experience by relying on the display of background neural or cognitive tendencies of the subject. Both paradigms are thus insensitive to the entirety of the phenomenon of decision making.

Historically, the perceptual judgement and value-based schools of thought in neuroeconomics have represented this division and have independently pursued the investigation of the same phenomena from different perspectives.[3] If both schools of neuroeconomic thought only capture or track a portion of the conscious experience of decision, their unification will be a necessary condition for holistic, fuller explanation of decision making.

Perception, Commitment, and Mechanisms of Choice

A canonical definition of a choice in decision neuroscience is the commitment to a proposition.[4] The field's aim is to discover how the brain makes this commitment through the accumulation, integration, and processing of evidence. Several models are prominent, including *drift diffusion*, in which evidence accumulation is understood as

[3] Glimcher 2014.
[4] Gold and Shadlen 2007.

accumulation of neural activity to a bound and a decision as crossing the threshold defined by the bound[5]. Unless evidence accumulates to a threshold, a choice isn't made, and this idea is reflected in neural activity. The usefulness of this picture provides a clear mechanism of choice—accumulation to a bound—and makes it easy to imagine and study conditions that could be manipulated to interfere with the cognitive and neural activity that constitutes evidence accumulation. An experimenter could decrease the time allotted to cross the threshold to demand a faster integration of incomplete evidence, or introduce noise into the stimulus to influence the time it takes for evidence to reach a decision threshold. On the drift diffusion model, commitment is constituted biologically by populations of neurons reaching or exceeding a firing rate threshold.

On the drift diffusion model, biases like value, confidence, or belief can manifest in different ways. The baseline firing rate of decision neurons might be elevated or the bounds lowered so that commitments happen faster. Intuitively, this kind of adjustment can be the result of learning or expertise. After through training in some domain, one does not need an excessive amount of evidence to commit to a decision. Consider an experienced surgeon compared to a first-year surgical resident. The expert operates with high confidence—they understand the high value targets in the world and aim for them. The expert surgeon can cut precisely, while the student would do well to measure twice and cut once. The expert requires less evidence to quickly reach the threshold.

[5] Gold and Shadlen 2007, Kiani and Shadlen 2009.

Another prominent model of neural commitment is the *race* model of decision making, which is an expanded version of drift diffusion.[6] Instead of a single decision threshold to surpass in one direction, both positive and negative evidence, that is, evidence toward available options, accumulate against one another over time. On this model, whichever option accumulates the most convincing evidence quickest "wins" and generates a commitment to that option. If we imagine the race model as a two-dimensional or dual aspect drift diffusion scheme, the same kinds of bias-like adjustments are still fair game, but this time in two directions. Bias toward a positive bound for instance could manifest as increased baseline neural activity in neurons associated with the perception of option A, while bias toward the negative bound might be constituted by the raising of the lower bound firing rate of neurons associated with the perception of option B. The effects could be compensatory, additive, or subtractive, but in any case, constitute the *flexibility* of decision making at the neural level.

Neural Bayesianism

Both the drift diffusion and race models of decision making are mathematically analyzable from a Bayesian perspective at the behavioral and neural levels—that is, as

[6] Gold and Shadlen 2007

an iterative, probabilistic process that conditions on priors.[7] The thought is that both behavior and neural activity follow patterns that don't start fresh with every experience. Rather, they are shaped and influenced by prior ones. Behaviorally, this might manifest as an agent's decision time decreasing over multiple trials. A decider can build *expectations* about a situation and thus perceive, reason, and act more efficiently when making choices in that situation. The Bayesian explanation of this phenomenon is that the agent conditionalizes on priors with each decision, increasing the probability that a certain choice is favored given certain conditions. The strength of Bayesianism is that it *links* behavioral and neural data and can predict both. Neural firing rates associated with different choices also appear to conditionalize on priors and fit behavioral data.

The Bayesian approach to decision making helps characterize the way that decision agents and neurons commit to a proposition *P,* which could also be understood as the way the agent takes the world to be. For instance, say I am part of a perceptual decision experiment. When I decide that there is a red square in front of me, I accrue perceptual evidence related to the proposition *that there is a red square* when I view the stimulus provided. That evidence accumulates to a threshold either in favor of *P* being true (+*P*) or *P* being false (–*P*) and is reflected in my neural activity, which on subsequent trials will change because of my prior experiences discriminating red squares. As time progresses, I might require less time to commit to the proposition that

[7] Lin 2024, Gold and Shadlen 2007.

there is a red square. An experimenter could manipulate the stimulus by desaturating or distorting the image so it's less clear to me whether there is in fact a red square, and should expect to see a difference in both my decision behavior and neural activity that, with learning, should conditionalize in the same way as the non-noisy condition.

The Bayesian picture of decision making provides a clean and uncluttered way to reason about how the brain decides by relying heavily on conditionalization and probability. It works well for perceptual decisions.[8] It also works well for value-based decision making and can explain why agents develop habits. In both cases, conditionalizing on priors with positive outcomes increases the probability that an agent would decide similarly given similar stimuli. However, these low stakes decisions may be too artificial to capture the more complex forms of decision making associated with things like moral or political reasoning, which entail a heterogeneous mix of perceptual and value judgements to produce a decision like voting or responding to ethical puzzles. It is this kind of decision making that needs to be better understood to get a better grip on human decision making and how it might go awry in neurological and psychiatric diseases, or how it might be optimized for policymaking.

[8] SHadlen and Hanks et al. 2006.

Fact and Value

If we are to understand the nature of human decision making, we must acknowledge that it is captured neither by simple perceptual discrimination nor by preference alone. These two kinds of judgement are entailed in every human choice—every decision is a mixture of fact and value. A decider intuits the world in perception and responds to it with action via a cognitive process that integrates objective facts about the world and the environment with the subjective values of the agent to produce a choice action. Decision making happens at the interface of fact and value, and choice is the expression of their integration. This integrative aspect of decision is what gets missed by the perceptual judgement and value-based experimental paradigms in decision neuroscience. Both camps seek to explain the same phenomenon from different, reciprocally impoverished angles that prioritize one element of a process with dual and equal aspects. If they are integrated, such a paradigm would be a much better way to track salient factors of decision making as it is experienced.

One way to study the relation between perceptual judgement and value-based aspects of decision making is through choice experiments with higher (imaginary) stakes than small monetary rewards or accumulating points. Reward-based paradigms introduce an inherent confusion by their design—they do not account for the subjective value represented by the choices themselves. There is a difference between choosing

something because it is valuable and choosing something because it will provide a reward, but neither perceptual judgement-based nor value-based experiments can tease out which is motivating choice behavior.

In perceptual judgement experiments, subjects could desire the reward and so decide in whatever pattern delivers reward, not *because* the choice is correct. The accuracy of judgements of motion coherence or the location of light flashes, for example, might be driven by reward point values (or in non-human primates, concentrations of sugar) rather than true perceptual discriminations. A subject might not fully understand the stimulus nor the choice they are making yet identify a pattern of actions that yields a reward. A subject might not *know* that the dots are moving to the left, yet may still choose in a way that looks like they do because that behavior yields a reward. In the same way, someone who cannot add might respond "four" when you ask them for the sum of two plus two because they have determined the appropriate sounds to make in the context of the question. We would not call the latter person a competent speaker, nor should we call the former subject a competent chooser. In these kinds of cases, choice behavior merely looks like fact-based decision making. Neural measurements taken during this kind of behavior would then not be correlates of decision making in the sense of deliberation-based action in a set of conditions, but some kind of masked reward-seeking behavior.

In value-based judgement, decider may choose by comparing arbitrary point or monetary values that, if switched or adjusted would produce different decisions that ignore other aspects of the stimulus. A subject might place an inordinate amount of value on some worthless product like a rubber ball instead of on a diamond ring simply because it is assigned an arbitrary value as part of the experimental conditions. Again, they can learn to play the game set up by a decision making experiment, which will yield data about how subjects play the game rather than about deliberative decision making-related cognition and action selection. If a subject abandons their sense of the value of objects because of the rules of an experimental game, the experiment does not reveal why or how people make the decisions, but rather tell us about how experimental test subjects play exquisitely designed perceptual and value games. They tell us even less about how deciders deal with the relation between fact and value, and so very little about how to explain the neural or cognitive basis of decision experiences or how they might be disrupted in pathological circumstances. To get at this aspect of decision making, a different kind of experiment is needed.

Complexity and Consequence

Simple perceptual or value-based decisions are *safe.* Tasks like the identification of shapes or the direction of the motion of dots on a screen, betting with small sums of

money or points, or choosing between consumer products all solicit decisions in the context of relatively low stakes. What makes these decisions low stakes are their low-impact consequences. Incorrect or inaccurate decisions about the character of visual experiences, poorly placed hypothetical bets, or choosing one brand of consumer product over another don't carry much existential or moral weight. Even in experiments with dehydrated non-human primates, the stakes involve a physiological need, but are still rather low, all things considered. While the perceptual and value-based paradigms produce insight about basic choices, they fall short of shedding light on some of the most natural questions about decision making: How do we make high stakes, complex decisions? Should I take out a mortgage? Marry this person? Have a child? How do we reason about what we *ought* we do? How *should* we decide in a given situation? How *do* we weigh options and select from alternative courses of action?

These *normative* questions are relevant to ordinary human life and contexts like policymaking, law, war, and medicine, where decisions affect the way the world is and how our lives, and the lives of others, will go. If a choice makes a deciders life better or worse, normative consequences bear on that decision. However, decision making experiments in decision neuroscience seem to ignore this fact, either focusing on categorizing perceptual reports or the idiosyncratic preferences of deciders and searching for their neural correlates via recording or imaging. In principle, neither approach fully captures the decision experience, but together, they might. A more

holistic approach to decision neuroscience could track the plurality of aspects that constitute the experience of decision making, and so lead to new discoveries.

Experimental Trolleys

Safe and simple perceptual or value-based decisions only go so far in revealing the nature of decision making. The stakes are low and choices are constrained by the rules of experimental games, producing insight about the accuracy of perceptual reporting ability or the variation in one's preferences as a function of arbitrarily assigned value. These insights are valuable for understanding attitude reports and economic agency, however they leave the richer phenomenon of deliberative, consequential choice employed by lawmakers, physicians, and ordinary people when they make serious, multidimensional decisions relatively unexplored. One reason for this is the *reactive* nature of the experimental setups. If choices are made in response to a briefly flashed or cost-based mathematical stimulus, subjects are processing a situation and making a choice that is constrained to the rules of the experimental game. However, decision making is *deliberative.* When we decide, we *think* before we act. (Or at least we should, ideally.) The reactive decisions elicited in classical experimental paradigms that I have been mentioning would be derided as hasty in the real world.

How could we probe the deliberative side of decision making? The first constraint is that such experiments would have to be done with human subjects. Non-human animals may deliberate, but we cannot know what it's like for them to do so, and therefore cannot reliably establish links between thought, neural activity, and action. Human subjects can provide verbal reports about their deliberative processes, and so reveal the motivations, reasons, and values that ground their decisions. Human subjects also permit the use of complex perceptual and narrative stimuli, thus allowing for more fine-grained condition setting and stimulus manipulation.

If we want to know something about the deliberative process, we need to know something about the role of *thought* in decision. Reactive perceptual and value-based judgements entail some basic propositional thought about, e.g. the presence or absence of a shape or the odds of a bet, but they do not entail complex deliberations anything like those that arise in the context of higher-stakes decisions. A policymaker might integrate moral, political, and economic thought into a decision, or a physician might integrate medical, ethical, and moral thought. To get a handle on the mechanisms of higher stakes decision, controlled experiments will have to be conscious of the different kinds of thought such decisions might entail.

One way to accomplish an experimental synthesis between the perceptual and value-based schools of neuroeconomic thought that is conscious of the cognitive context set my experimental subjects' background values and the multimodal nature of higher

stakes decision making is by observing subjects' choice behavior in response to moral dilemmas. Moral dilemmas encapsulate many, if not all, of the aspects that might be entailed in a decision—perception, morality, value, cost, benefit, well-being, consequence, and action. They are complex stimuli with the power to launch a subject into deeply deliberative thought about what choice to make, and have the added bonus of eliciting people's deeply held background values. Further, there are some studies that have probed the neural dimension of decision making in the context of moral dilemmas, but the studies have either aged or are largely theoretical in nature.

Consider the following experimental setup designed to probe reported and neural differences across decision makers as a function of modified versions of the notorious Trolley Problem:

Experimental Design

- - - - - - - - - - - - - - - - - - - *Experiential Stream* -

Value Questionnaire → Trolley Problems → Debriefing Questionnaire

Baseline fMRI fMRI monitoring Debriefing fMRI

- - - - - - - - - - - - - - - - - - - *Neural Stream* -

The experiential stream of the experiment is designed to collect attitude reports and internal factors about the decider (Value Questionnaire), present external stimuli and an option set simulating a high-stakes decision (Trolley Problem), then collect reflections and justifications about decision behavior (Debriefing Questionnaire). All the while, the neural stream of the experiment deploys fMRI scanning to track baseline brain activity, and activity during decision making about one's values and propensities, the Trolley Problem, and reflections on these decisions.

The Value Questionnaire is designed to probe subjects' perceptions of their own values, in terms of their risk tolerance and basic conception of the morality of actions. The idea is that both questions will yield attitude report data that sets the context for the subjects' decisions and provides some evidence about the perspective from which they are making choices. Here it is:

Value Questionnaire

1. Do have a high, medium, or low tolerance for risk?
    a. High tolerance for risk
    b. Medium tolerance
    c. Low tolerance
    d.

2. Do you think that some actions are always wrong or that the wrongness of an action depends on the situation?
    a. Some things are always wrong.
    b. It depends on the situation.

The control group would be presented with the standard version of the Trolley Problem, as outlined below:

Switch Trolley

A trolley is speeding down some tracks in your immediate vicinity toward a fork. On the main track of the fork are five people who have been unwillingly tied to the tracks in danger of being run over. On the side track of the fork, there is one person unwillingly tied down. Within your reach, there is a switch that can divert the trolley from its path on the main track to the side track. If you do not flip the switch, the trolley will continue on the main track and kill the five people tied down there. If you flip the switch, the one person on the side track will die, but the five will be saved.

The Trolley Problem and its variants entail all the elements I have argued are involved in the decision making experience: background value (internal factors about the decider), perceptual judgement (understanding the situation and its elements, value-based judgement (ethics-influenced action options), and an action decision. The Trolley Problem unifies the aspects of decision making explored by the different schools of decision neuroscience, and we can adjust it to do so even more explicitly, using only language modifications to manipulate internal and external decision factors.

One experimental group can be shown the following modified trolley case designed to probe the influence of perceptual judgement on decision (changes to the original case are denoted by italics):

Storm Trolley

During a blinding snowstorm, a trolley is speeding down some tracks in your immediate vicinity toward a fork. On the main track of the fork are *what looks to you* to be five people who have been unwillingly tied to the tracks in danger of being run over. *However, since the snow is so blinding, you cannot be sure what exactly is obstructing the track.* On the side track of the fork, there is *what looks like* one person unwillingly tied down. *But again, due to snow blindness, you cannot be certain.* Within your reach, there is a switch that can divert the trolley from its

path on the main track to the side track. If you do not flip the switch, the trolley will continue on the main track and *collide with the obstruction on the main track.* If you flip the switch, the *train will collide with the smaller obstruction on the side track, and the larger one will be spared.*

Storm Trolley introduces *perceptual noise* into the situation. Unlike Switch Trolley, an element of uncertainty about what the consequences of the agent's choice will be enters the decision deliberation. Another experimental group would be presented with a case incorporating value-based noise:

Bribe Trolley

A trolley is speeding down some tracks in your immediate vicinity toward a fork. On the main track of the fork are five people who have been unwillingly tied to the tracks in danger of being run over. On the side track of the fork, there is one person unwillingly tied down. Within your reach, there is a switch that can divert the trolley from its path on the main track to the side track. If you do not flip the switch, the trolley will continue on the main track and kill the five people tied down there. If you flip the switch, the one person on the side track will die, but the five will be saved. *However, someone makes you the following offer as the train*

speeds down the tracks: If you let the trolley continue on the main track, he will give you $10,000, and if you decide to switch the trolley, you will get nothing.

Bribe Trolley introduces *value-based noise*, thereby putting pressure on the balance between perceptual and value-based judgements in the opposite direction of Storm Trolley. With these cases laid out as stimuli, we have a neutral case (Switch Trolley) that unifies decision variables, plus two experimental cases that represent perceptual and value-based distortions. One final case can introduce a moral variable:

Special Obligation Trolley

A trolley is speeding down some tracks in your immediate vicinity toward a fork. On the main track of the fork are five people who have been unwillingly tied to the tracks in danger of being run over. On the side track of the fork, *the person you love the most* is unwillingly tied down. Within your reach, there is a switch that can divert the trolley from its path on the main track to the side track. If you do not flip the switch, the trolley will continue on the main track and kill the five people tied down there. If you flip the switch, *your loved one* on the side track will die, but the five will be saved.

The motivation behind the Special Obligation trolley is to introduce moral noise. In the same way that perceptual and value-based noise can influence decision, moral noise can too. Presumably the special obligation will disrupt the decision experience in significant ways, hence its usefulness for probing the nature of choice. The final wrinkle of these experimental trolley cases is a time condition. The cases themselves have e an element of urgency in their describing a speeding trolley, however the time allotted to make the decision might also play a significant role in the way decisions are made. As such, there would be a Short Window and a Long Window within each group, with the former allotting 1 minute to make the decision and the latter allotting 3 minutes.

The last component of the experiment I am proposing would be a debriefing questionnaire designed to capture subjects' justifications for their decisions. The idea is to collect attitude reports about the decision experience and examine their relation to pre-decision reports and actual decision behavior. They would be:

Debriefing Questionnaire

1. What played the most prominent role in the reasoning behind your choice?
    a. The details and facts of the situation
    b. Your own personal values, beliefs, and feelings

2. What was your level of confidence in your decision?
    a. Very confident
    b. Fairly confident
    c. Not confident

What could we learn from these experiments? In line with the experiential and neural streams of the experiment, the hypotheses for the experiment are as follows:

Hypothesis 1

Perceptual, value-based and moral noise will induce decisions that contravene reported values and attitudes, and decisions made in a non-noisy context.

Hypothesis 2

Trolley problems will engage brain areas in the anterior cingulate cortex associated with processing conflict, the dorsolateral prefrontal cortex associated with abstract reasoning, and the ventromedial prefrontal cortex associated with deontological judgements[9].

[9] Greene 2004, Joon 2016, and Kreighoff 2011.

Hypothesis 3

The introduction of noise into the decision scenario will have dual effects on the experience of decision making that manifest as increased decision (reaction) time, and reduction in activity levels in those areas resulting from uncertainty.

Hypothesis 1 is a conjecture about the distorting effects that internal and external noise can have on the decision process, Hypothesis 2 is meant to be a connectivist-style conjecture about the simultaneous recruitment of brain areas in complex decisions. Finally, Hypothesis 3 is a conjecture about the effects of noise on the accumulation and consideration of evidence, and its translation to effects on decision making behavior and neural activity.

For the Value Questionnaire, I would expect risk tolerance reports to be balanced across subjects. Risk tolerance is personal and varies greatly across individuals. I would also expect a roughly even split between people's beliefs about the morality of actions being principled and inflexible or context dependent. The data this experiment might yield might be as follows:

Value Questionnaire

Risk	**Risk Seeking**	**Risk Neutral**	**Risk Averse**
	33%	33%	33%

Morality	**Always Wrong**	**Situation Dependent**
	50%	50%

For the Trolley Problems, I would expect that in the Switch case, subjects will find it wrong to switch, in line with previous findings in psychology and neuroscience experiments. I would further expect there to be a wider gap between Switch and No Switch choices in the Short Window condition, given the tendency of deliberation time to affect decision. For the noise conditions, I would expect that noise reduces the gap between Switch and No Switch responses across time windows because the choices become less clear. In Storm, I suspect perceptual noise would decrease the switch rate in both the short and long window conditions since the quality of the evidence is decreased and uncertainty is introduced. In Bribe, I expect that the lucrative offer and the estrangement between the subject and the potential victims would increase rates of the No Switch response across time windows, but that the No Switch response would be reduced in the Long Window condition because subjects have more time to

deliberate and rail against coercion. In the Special Obligation case, I expect a sharp increase in Switch responses across both time windows, and an even more drastic signal for Switch in the Long Window condition, due to the combined effect of extended deliberation and emotionally significant attachment to the lone individual. This data might look something like the following:

Trolley Problems

	Short Window		*Long window*	
	Switch	**No Switch**	**Switch**	**No Switch**
Switch (Control)	80%	20%	70%	30%
Storm (Perceptual)	60%	40%	55%	45%
Bribe (Value-based)	55%	45%	70%	30%
Special Ob. (Moral)	10%	90%	5%	95%

For the Debriefing Questionnaire, I would generally expect the Short Window condition to produce more value-centric decisions due to shorter decision time, and expect subjects to rely on their "gut" more than the objective facts of the situation since

there is a short time to collect evidence and deliberate. In the long window condition, I would expect facts about the situation to play a more prominent role in deliberation and decision making, given that a longer time window allows for deeper consideration of the dynamics of the situation. For confidence levels, I would expect Short Window decisions to have predominantly higher confidence due to the pressure applied by the time constraint. In the Long Window condition, I would expect confidence levels to cluster at medium levels, since increased deliberation time introduces opportunity for subjects to question their intuitive reactions. Here is a data projection:

Debriefing Questionnaire

	Short Window			***Long window***		
Prominent Factor	**Fact**		**Value**	**Fact**		**Value**
	30%		70%	40%		60%
Confidence Level	**High**	**Medium**	**Low**	**High**	**Medium**	**Low**
	50%	30%	20%	40%	40%	20%

Neuroimaging results are hard to project, especially given a complex stimulus like the Trolley Problem and the variants spun up here. However, decision making is a complex phenomenon, so we would do well to at least try. In a cursory review of

relevant literature on moral cognition, the anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), and ventromedial prefrontal cortex (VMPFC) feature prominently for their roles in conflict reasoning, abstraction, and utilitarian or deontological judgements, respectively. As a result of these findings and the functional roles of these brain regions, I would expect them to be active during engagement with Trolley Problems, so I will speculate about how the different conditions framed above might affect activity in those regions. Here is a neural data projection:

Neuroimaging

Anterior cingulate cortex (ACC)	**Switch**	**Storm**	**Bribe**	**Special**
	Active	Elevated	Elevated	Markedly Elevated

Dorsolateral prefrontal cortex (DLPFC)	**Switch**	**Storm**	**Bribe**	**Special**
	Active	Elevated	Reduced	Markedly Reduced

Ventromedial prefrontal cortex (VMPFC)	**Switch**	**Storm**	**Bribe**	**Special**
	Active	Reduced	Elevated	Reduced

Importantly, I think the question isn't *which* of these brain areas will be active across the different Trolley conditions, but *how active they will be.* I am projecting that these regions form some kind of network that will all be active in response to the Trolley stimuli, and will modulate in the presence of different variations. That is, I think decision making would be revealed to be heavily multi-regional in nature and involve these regions. I would expect ACC activity to be elevated in all Trolley variations as they all raise a moral conflict, but markedly elevated in the Special Obligation case because the introduction of an endangered significant other elevates the conflict. I expect DLPFC activity to be elevated in the storm condition because perceptual noise forces the agent to use an additional layer of counterfactual reasoning to think about how they might act under increased uncertainty, while DLPFC activity could be reduced in Bribe because a valuable reward is readily available, and further reduced in Special Obligation because the presence of the significant other makes the decision much more real and consequential despite still being hypothetical. Lastly, I would expect the VMPFC, since it has been associated with reasoning about utility, to be elevated in Bribe again due to the opportunity for a big reward albeit it at a serious cost, and reduced in Storm and Special Obligation because these kinds of decisions are less about utility and more about uncertainty and saving a significant other, respectively.

One last consideration is the relation between the allotted decision time on brain activity. My projection is that in the Short Window conditions, brain activity might be

more intense for shorter durations, since deciders have a limited time to accrue evidence to a decision bound. In the Long Window condition, I might expect for activity to be less intense but persist over longer periods. There might also be interconnected fluctuations in brain activity in these regions as time passes in the longer time window and the subject continues to deliberate. It would be interesting to see if verbal reports about the decision experience might be taken from the subjects and compared with changes in neural activity over time.

These neuroimaging projections are admittedly highly speculative. Nevertheless, it would be revealing to know if and how the ACC, DLPFC, and VMPFC interact in the context of the complex decisions presented by the Trolley stimuli. In any case, my expectation would *not* be that any one of these areas is inactive during the kind of decision making that this experiment elicits.

Gaps and Possible Improvements

This discussion is highly speculative and theoretical and the same goes for the experiment proposed, however I would like to offer some points about what would need to be refined or further considered in realizing the approach discussed this paper. Since I have already discussed these matters at length, I'll only mention these

shortcomings briefly and sketch what might be needed to address them, without developing the theory or experiment further at this time.

First, the definition of value is still hard to pin down. Defining value is philosophically complex, so additional work would need to be done to sharpen the operative notion in this context. There are multiple conceptions of value in the philosophical literature, e.g. desire/preference satisfaction, welfare, moral value, so it could be helpful to choose one and calibrate the theory and experiment to it further. Second, I have specified to use fMRI in the experiment for investigating neural activity. This seemed intuitive to me, but another approach is to use EEG. This would entail deeper exploration of what signals are associated with the brain areas I've focused on and how they might be related to different aspects of decision making. Third, the sample size here is left undefined. We could set it a N=100, however some further thought would be needed to understand what sample size would be needed to resolve significant neural and behavioral differences. Lastly, the questionnaires are quite brief. More research would be needed to determine additional helpful questions.

Conclusion

I hope to have conveyed that human decision making is not a simple process, and so leaves stones unturned when studied in a simplistic way. Understanding the

cognitive, behavioral, and neural mechanisms of such a complex process is not easy, but that does not mean it is impossible. Continued disjointed investigation of perceptual and value-based choice will necessarily miss certain aspects of the phenomenon, and it is only through an attempt at unification that serious progress might be made.

References

Glimcher, P. (2014). Value-Based Decision Making. *Neuroeconomics* 2nd Edition.

Glimcher, P. (2022). Efficiently Irrational: Deciphering The Riddle of Human Choice. *Trends in Cognitive Sciences* 26(8): 669-687

Glimcher, P. and Tymula, A. A. (2023). Expected Subjective Value Theory (ESVT): A Representation of Decision Under Risk and Certainty. *Journal of Economic Behavior and Organization* 207: 110-128.

Gold, J. I. and Shadlen M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience* 30: 535-574.

Gothard T. and J. Davies (2204) A Spiking Neiron Model of Moral Judgement in Trolley Problems. *Scientific Reports* 14: 21733.

Greene, J., L. E. Nystrom, A. D. Engell, J. M. st al (2004). The Neural Basis of Cognitive Conflict and Control in Moral Judgment.

Hume, D. (1739). *A Treatise of Human Nature* (E. C. Mossner, Ed.). 2005. Penguin.

Joon, W.H., K. Prehn, Z. Fang, et al (2016). Moral Competence and Brain Imaging: A Resting State fMRI Study. *Neuroimage* 141: 408-415.

Kiani, R. and Shadlen, M. (2009). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. *Science* 324 (5928): 759-764

Kreighoff, V., F. Waszack, W. Prinz, and M. Brass (2011). Neural And Behavioral Correlates of Intentional Actions. *Neiropsychologia* 49 (5): 767-776.

Levy, D. J. and Glimcher, P. (2012). The Root of All Value: A Common Neural Currency For Choice. *Current Opinion in Neurobiology* 22(6): 1027-1038.

Lin, H. (2024). Bayesian Epistemology. *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.).

Nguyen D., E. L Rich,., J. D. Wallis et al. (2025). Neural Representation of the Decisional Reference Point in Monkeys. *Proceedings of the National Academy of Sciences U.S.A.* 122 (48) e2514110122.

Shadlen M. N., T.D. Hanks, A.K. Churchland, et al (2006). The Speed and Accuracy of a Simple Perceptual Decision: A Mathematical Primer. In *Bayesian Brain: Probabilistic Approaches to Neural Coding*, ed. K Doya, S Ishii, R Rao, A Pouget. MIT Press.

Unger, P. K. (1996) *Living High and Letting Die.* Oxford University Press. New York.