Identity, information and situations

Published version:

DOI:10.1093/jeea/jvad033

(Article begins on next page)

Identity, information and situations

Daniele Pennesi
ESOMAS Department and Collegio
Carlo Alberto, University of Torino

April 2023

Abstract
This paper introduces a model of individual behavior based on identity, a person's sense of self. The individual evaluates situations, i.e., sets of available actions given a belief about the actions' uncertain payoffs. In some situations, a psychological cost arises because the individual's identity prescribes an action that differs from the one maximizing material benefits. The model shows that a common process of weighing psychological costs and material benefits drives the choice of both information and future opportunities. As a result, information avoidance is akin to preferring fewer opportunities, such as crossing the street to avoid a fundraiser. The model provides a coherent rationalization for diverse behaviors, including willful ignorance, opting out of social dilemmas, and excess entry into competitive environments. The psychological cost varies non-monotonically with the quality of information or with having more opportunities. Non-monotonicity complicates the identification of prescriptions from behavior, a difficulty that is partially resolvable by observing specific choices. (JEL: D01, D83, D91)

Keywords: Identity, Self-image, Information Aversion, Willful Ignorance, Gender Identity, Preference for Commitment.

1. Introduction

The desire to protect one's identity, a person's sense of self, is a well-known determinant of behavior (Akerlof and Kranton, 2000; Bénabou and Tirole, 2011). Identity-conscious individuals consider the material consequences of their actions as well as how these actions relate to "who they are." A costly *identity trade-off* emerges when identity-based behavior conflicts with the actions that maximize material payoffs. To reduce this trade-off, individuals

Acknowledgments: I thank Vasiliki Skreta and three anonymous referees for their invaluable suggestions. I am particularly grateful to Andrea Gallice and Ignacio Monzón for their helpful comments. I also benefited from discussions with Larbi Alaoui, Botond Kőszegi, Antonio Penta, Aron Szekely, seminar participants at the SAET 2019 (Ischia), D-TEA 2021 (Paris), BRIC 2022 (Prague), XVI GRASS (Ancona), Collegio Carlo Alberto, and Bologna. A previous version of this paper circulated under the title "Identity and information acquisition."

E-mail: daniele.pennesi@unito.it

adopt sub-optimal behaviors, such as avoiding instrumental information (Dana et al., 2007), or restricting future opportunities (e.g., crossing the street to avoid a fundraiser, see Andreoni et al., 2017; DellaVigna et al., 2012).

Existing models of identity-based behavior study special cases of information avoidance or special cases of restricting future opportunities, and their results depend on simplified settings and specific assumptions, such as self-uncertainty (e.g., Bénabou and Tirole, 2011; Grossman and van der Weele, 2016), which limit their applicability to complex decisions.

In this paper, I present a unified model[1] of how identity influences the selection of information and future opportunities. My model provides a common framework that organizes existing but scattered theoretical results and generates novel insights that can benefit both applied and theoretical research on identity. To achieve this, I assume that the individual evaluates *epistemic situations*: pairs (F, q), where F is a set of available actions with uncertain payoffs and q a belief about the states of the world. In each epistemic situation, the individual's identity prescribes an action in F, and another action maximizes the individual's material payoffs. When these two actions differ, the identity trade-off emerges.

My first contribution is to establish that there is a common process describing the choices of information and of future opportunities. Acquiring information leads to greater material payoffs, but may also come with an increased psychological cost. An analogous cost-benefit analysis applies to the choice of future opportunities. As a result, information avoidance is akin to a preference for commitment, while a "demand for beliefs" is analogous to a demand for non-instrumental flexibility (i.e., a preference for including materially inferior options). Therefore, my model provides a coherent rationalization for disparate behaviors, including willful ignorance, opting out of social dilemmas, and excess entry into competitive environments.

1. More precisely, I develop a class of models which all have the "material payoffs minus psychological cost" structure. Parameterizations of the cost identify models within the class.

My second contribution is to show that the cost of information (and of flexibility) may respond non-monotonically to the quality of the information (and to more flexibility). This non-monotonicity is important when deciding how much information to disclose to an identity-conscious receiver, and distinguishes this model of costly information acquisition from others.

My third contribution is to show that by observing choices over specific epistemic situations, it is possible—although notoriously difficult— to partially identify the unobservable identity of an individual. Therefore, my model provides new tools to inform laboratory and field experiments studying identity.

To illustrate my approach, consider the well-known "moral wiggle room" experiment (Dana et al., 2007).[2] In this variation of the dictator game, ω_1 and ω_2 are two equally likely states of the world. The dictator can choose between two actions a, b. The state-contingent payoff (x, y) of each action represents a monetary allocation (x for the dictator and y for the recipient). Table 1 shows the payoffs in Dana et al. (2007). When the dictators know that the state is ω_2,

Table 1. Actions and payoffs in Dana et al. (2007).

	ω_1	ω_2
a	$(6, 5)$	$(6, 1)$
b	$(5, 1)$	$(5, 5)$

74% of them play b. In the main treatment, the dictators do not know the state but can learn it at no cost. In this case, only 56% of the dictators decide to learn it, and all those who choose to remain ignorant play a. The results support the intuition that ignorance allows dictators to act selfishly while preserving an altruistic self-image.

Epistemic situations help to describe the moral wiggle room experiment and to rationalize the observed behavior. Under ignorance, the dictator is in the epistemic situation $(a \cup b, \hat{p})$, where $\hat{p}(\omega_1) = \hat{p}(\omega_2) = 0.5$ is the prior. In this

2. See also Larson and Capra (2009), Matthey and Regner (2011), Grossman (2014), Feiler (2014), Grossman and van der Weele (2016), and Spiekermann and Weiss (2016).

case, a is the action with the highest material payoff and is also the prescription (for an altruistic individual). Indeed, the payoffs of a and b are ex-ante the same for the recipient. Thus, there is no identity trade-off under ignorance. Similarly, there is no trade-off if the state is ω_1, corresponding to $(a \cup b, \delta_{\omega_1})$, because a is, again, the payoff-maximizing action and the prescription. The identity trade-off emerges in $(a \cup b, \delta_{\omega_2})$, where the prescription (of an altruist) becomes b, but the payoff-maximizing action is a. Willful ignorance enables the dictator to avoid the identity trade-off and it corresponds to a strict preference for the epistemic situation $(a \cup b, \hat{p})$ over a "lottery" that yields the epistemic situations $(a \cup b, \delta_{\omega_1})$ or $(a \cup b, \delta_{\omega_2})$ with equal probability. Epistemic situations capture an alternative way of eluding the identity trade-off: restricting future opportunities. A reluctant altruist can strictly prefer committing to a, rather than having the flexibility of $a \cup b$.

The moral wiggle room illustrates a particular case of the cost-benefit analysis that drives information acquisition in my model. Acquiring information (weakly) increases the expected material payoffs, but also modifies the posterior beliefs, and thus the prescriptions, potentially exacerbating the identity trade-off. This analysis explains willful ignorance even when the identity trade-off is not evident (as is the case of poorly informed donors discussed in Section 3.2), and in domains other than social dilemmas, such as health. For example, a routine medical test recommending a change in behavior, such as consuming less red meat, generates the identity trade-off for a stereotypical masculine identity. This helps explain why males engage less than females in preventive healthcare (Courtenay, 2000). However, ignorance is not always motivated by material gains: through its indirect effect on the prescriptions, even non-instrumental information is valuable or costly, generating a demand for beliefs. A religious person may prefer not to know if a life-saving medicine contains prohibited substances. Information is non-instrumental because the optimal behavior is to take the medicine anyway, whereas learning that it contains prohibited substances could be extremely costly, making ignorance optimal.

The psychological cost of information may respond non-monotonically to the quality of information. Non-monotonicity has multiple consequences. For instance, a donor may prefer to remain ignorant rather than learn that their preferred charity may be low quality (Niehaus, 2014). At the same time, they would acquire information if it eliminated such uncertainty (see Section 3.2). Non-monotonicity thus complicates the charity's decision about how much information to disclose to potential donors and also generates an asymmetry in the interpretation of choice data. Observing information avoidance may be informative about identity concerns, but observing information acquisition may not. This issue is addressed in Section 5, where I show that it is possible to infer prescriptions from the choice of information when it has no material value. Intuitively, in the absence of material gains, any information preference necessarily indicates a variation in the prescriptions.

Similarly to acquiring information, choosing future opportunities resolves a cost-benefit analysis. More flexibility increases the material payoffs, but also modifies the prescriptions, thus it may alter the identity trade-off. Therefore, having fewer opportunities is sometimes optimal. A preference for commitment explains why a reluctant altruist may prefer to escape from a situation where they could act prosocially (e.g., DellaVigna et al., 2012; Andreoni et al., 2017; Schwartz et al., 2021). It also rationalizes other identity-protective behaviors including "flexibility stigma" (the under-use of flexible work arrangements by fathers, Williams et al., 2013), and excess entry into competitive environments. If the prescription is to compete, entering the competition is the only way to eliminate the identity trade-off, even if the individual knows that doing so is materially sub-optimal. Therefore, my model gives an identity-driven explanation for why some entrepreneurs enter a market even if the investment has a negative net present value (NPV), or why males compete more than females (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011). The identity-driven explanation does not require overconfidence to rationalize excess entry. Commitment, moreover, is the only possible strategy to reduce the identity trade-off in the absence of uncertainty (e.g., Dana et al., 2006).

In parallel with a demand for non-instrumental information, my model predicts a preference for non-instrumental flexibility (or commitment). This refers to a desire to include (or exclude) opportunities, even if they are never materially optimal. Section 5 shows how to exploit preferences for non-instrumental flexibility to partially identify the unobservable identity of the decision maker. I conclude the paper by extending the model to account for meta-prescriptions, such as prescriptions about the appropriate attitude towards information or future opportunities (e.g., Austen-Smith and Fryer Jr, 2005; Bertrand et al., 2015).

The paper is structured as follows: Section 2 introduces the model; Section 3 applies it to information acquisition; and Section 4 studies preferences towards situations. Section 5 studies the use of revealed choices to infer prescriptions; Section 6 discusses meta-prescriptions; and Section 7 contains the literature review. Appendix A contains all the proof; Appendix B, the additional material; and the online Appendix C, special cases of the results in the main text.

2. The model

Actions and Identity. There is a finite number of states of the world $\omega \in \Omega$. Actions f are functions from the states to the payoffs. Thus, $f(\omega)$ is the payoff of the action f in the state ω. The individual self-categorizes as a member of a social category and[3] once categorized, internalizes the prescriptions. These "*indicate the behavior appropriate for people in different social categories in different* ***situations*** *[emphasis added].*" (Akerlof and Kranton, 2000). I model situations as pairs (F, q), where F is a finite set of actions (a menu) and $q \in \Delta\Omega$ a belief over the states of the world, and call them *epistemic situations*. In each epistemic situation (F, q), the identity prescribes an action in F, called the

3. I do not distinguish between personal and social identity. The former concerns the "role" that an individual occupies (or believe they occupy) in a society (Stets and Burke, 2000); the latter focuses on "belonging" to a social category (Abrams and Hogg, 2006). The two notions differ more in terminology than in substance (Stets and Burke, 2000).

q-belief prescription in F and denoted by $f_{F,q}$. If the prescription is insensitive to beliefs, I call it an *absolute* prescription.

Each action in F determines a possibly empty set of beliefs under which it is the prescription. I assume that these sets are *convex*. To illustrate the convexity assumption, consider the moral wiggle room example. If a is the prescription in state ω_1 and also according to the prior, convexity implies that a is the prescription for all beliefs assigning a probability of at least 0.5 to ω_1.

The identity trade-off. The individual has a utility u over the payoffs. The material value of an action f in the epistemic situation (F, q) is its expected utility $\mathbb{E}_q[u(f)]$. In each epistemic situation, there is (at least) one action, denoted by $f^*_{F,q}$, that maximizes expected material payoffs.

The identity trade-off emerges in an epistemic situation (F, q) when $f_{F,q} \neq f^*_{F,q}$. Given a menu F, I define the *trade-off regions* as the (sets of) beliefs in which the identity trade-off is present. By the convexity assumption, trade-off regions are convex sets.[4] If there are n actions in F, there are at most $n(n-1)$ trade-off regions (see Proposition A.1 in Appendix A). Figure 1 illustrates a possible representation of the identity trade-off in the moral wiggle room, assuming $u(x, y) = x$ and that the prescription is b whenever the probability of state ω_2 is larger than 0.75.[5]

The value of epistemic situations. I assume that the individual evaluates epistemic situations by weighing material payoffs and the psychological cost.[6]

4. If q and q' belong to a trade-off region in which f is the payoff-maximizing action and g the prescription, all beliefs $\alpha q + (1-\alpha)q'$ for $\alpha \in [0, 1]$ will belong to the same trade-off region (see Section A.1 in Appendix A).

5. The choice of 0.75 is inconsequential. What matters is the identity trade-off at $q(\omega_2) = 1$.

6. A possible interpretation, consistent with empirical evidence, is that the individual expects their selection *from* the menu to be $f^*_{F,q}$. As with any two-period model, it is a prediction about future behavior, but the actual second-period choice may be different. In Section B.2 of Appendix B, I introduce uncertainty about the anticipated second-period choice.

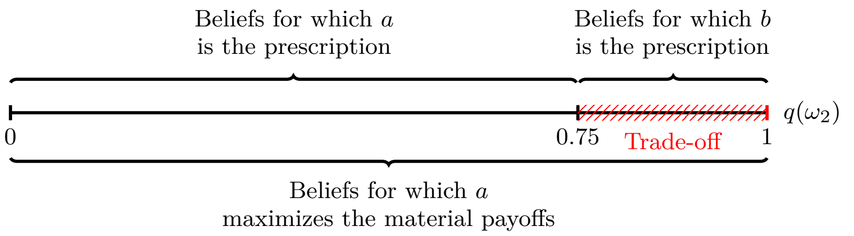

FIGURE 1. The black solid line is the probability of state ω_2. The prescription is a for all beliefs assigning a probability of less than 0.75 to state ω_2 (i.e., $f_{a\cup b,q} = a$ for all q with $q(\omega_2) \leq 0.75$), otherwise the prescription is b. The action a maximizes the material payoffs for all beliefs (i.e., $f^*_{a\cup b,q} = a$ for all q). The red pattern highlights the trade-off region: all beliefs for which the prescription and the payoff-maximizing action differ.

The value of (F, q) is

$$v(F, q) = \mathbb{E}_q[u(f^*_{F,q})] - d(f^*_{F,q}, f_{F,q}, q). \tag{1}$$

The function d is positive and satisfies $d(f, f, q) = 0$ for all actions and beliefs, meaning that there is no cost when there is no identity trade-off. If more than one action maximizes the material payoffs, I assume that the individual chooses the one that minimizes the psychological cost. One example of cost function is the discrete cost:

$$d_\kappa(f^*_{F,q}, f_{F,q}, q) = \begin{cases} \kappa & \text{if } f^*_{F,q} \neq f_{F,q} \\ 0 & \text{if } f^*_{F,q} = f_{F,q}, \end{cases} \tag{d_κ}$$

for all beliefs and a $\kappa \in [0, \infty]$ (see Gilboa et al., 2022). A different example is

$$d_e(f^*_{F,q}, f_{F,q}, q) = \varphi\left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})]\right), \tag{d_e}$$

for a convex, increasing and continuous function φ with $\varphi(0) = 0$ (e.g., Konow, 2000; Spiekermann and Weiss, 2016). It is as if the identity prescribes a utility level $\mathbb{E}_q[u(f_{F,q})]$ rather than an action, and obtaining more utility than prescribed is costly. The interpretation of the psychological cost is flexible; it can measure cognitive dissonance (e.g., when $d = d_e$) or the cost of a negative self-signal (à la Grossman and van der Weele, 2016). Figure 2 illustrates the moral wiggle room based on the assumptions of Figure 1 and with $d = d_2$.

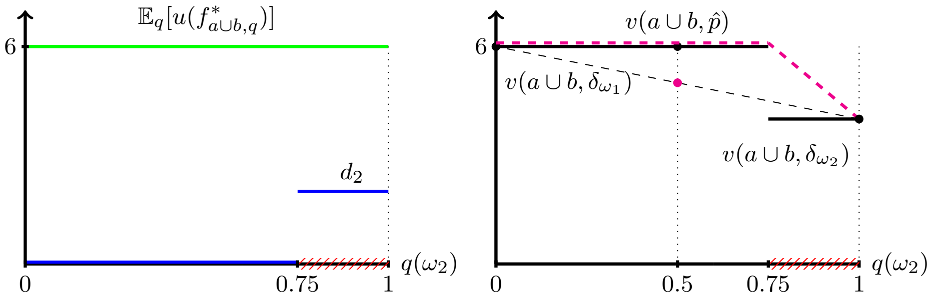

FIGURE 2. The moral wiggle room. Left panel: the material value of $a \cup b$ as a function of q under the assumption $u(x, y) = x$, so that $\mathbb{E}_q[u(f^*_{a\cup b,q})] = \mathbb{E}_q[u(a)] = 6$ for all q (green line). The cost d_2 (blue line). Right panel: the function $v(a \cup b, \cdot)$, given by the difference between the green and the blue lines of the left panel (black solid line). The smallest concave function that is greater than $v(a \cup b, \cdot)$ (dashed purple line). The value $v(a \cup b, \hat{p})$ (black dot) and the 1/2-1/2 average of $v(a \cup b, \delta_{\omega_1})$ and $v(a \cup b, \delta_{\omega_2})$ (purple dot).

Discussion of the assumptions. Given the focus of the paper, I do not explicitly model how identity, and thus prescriptions, emerge.[7] This assumption does make my model dependent on exogenous prescriptions, but the "degrees of freedom" are limited. First, the prescription is always an action in the menu, so the size of the menu limits the number of potential prescriptions. Moreover, the upper bound to the number of trade-off regions in Proposition A.1 and the convexity assumption further reduce the complexity of the identity trade-off. Lastly, there is no need to specify the prescription for all beliefs, but only for the prior and the posteriors (see, e.g., the moral wiggle room).

Second, the distinction between personal and social identity matters (see Footnote 3). *Personal norms* and *social norms* are distinct (see Bašić and Verrina, 2021) and which are relevant in a situation depends on contingent factors, such as observability by third parties. Social norms are "collective perceptions, among members of a population, regarding the appropriateness of

7. Identity is multidimensional and different identities of the same individual may be more or less ready to be activated in a situation. To discipline my model, I limit prescriptions to depend only on the epistemic situation and assume that self-deception is free, since the individual is free to select the most "convenient" identity. In the terminology of Kranton (2016), this is a model of short-run identity, where prescriptions and social categories are given.

different behaviors." (Krupka and Weber, 2013). Under this interpretation, $f_{F,q}$ is the behavior of the *prototypical* member of the population in the epistemic situation (F, q). Personal norms arise from "seeing the self in terms of the role as embodied in the identity standard" (Stets and Burke, 2000). Under this interpretation, $f_{F,q}$ is the action that meets the role's standards in the situation (F, q). Therefore, the information available to the analyst about the social or personal aspects of a choice restricts the potential prescriptions. Lastly, even if the analyst has no information about identity and situational aspects, the results of Section 5 help them to partially identify prescriptions from choices.

3. Information choices and non-monotonicity

3.1. Information avoidance

The individual has a prior $\hat{p} \in \Delta\Omega$ and information is an exogenously given *Bayesian experiment* μ consistent with $\hat{p}$. This is a probability distribution over beliefs that specifies the likelihood of deriving each posterior by Bayesian updating the prior. The experiment μ satisfies the consistency property $\hat{p} = \int_{\Delta\Omega} q d\mu(q)$, meaning that the expected information coincides with the prior. To avoid technicalities, I assume that μ has finite support. The implicit dynamic of the decision process is as follows: the individual selects a menu and acquires information μ or remains ignorant, receives information (if any), forms a posterior, and then selects an action from the menu. Lastly, the payoff materializes.

Given a menu F, its value under ignorance is $v(F, \hat{p})$. Information acquisition, instead, generates a lottery over posteriors. In this case, I assume that the ex-ante value of F is

$$V(F|\mu) = \int_{\Delta\Omega} v(F, q) d\mu(q), \tag{2}$$

corresponding to the average value of the epistemic situations associated with each posterior. I will refer to $V(F|\mu)$ and $v(F, \hat{p})$ as the *Identity model.* To emphasize the cost-benefit decomposition of information acquisition, I rewrite Equation (2) as

$$V(F|\mu) = \underbrace{\int_{\Delta\Omega} \mathbb{E}_q[u(f^*_{F,q})]d\mu(q)}_{W(\mu,F)} - \underbrace{\int_{\Delta\Omega} d(f^*_{F,q}, f_{F,q}, q)d\mu(q)}_{I(\mu,F)}.$$

The term $W(\mu, F)$ is the expected material payoff of F (e.g., Dillenberger et al., 2014), and $I(\mu, F)$ the average psychological cost. Information always has a positive material value ($W(\mu, F)$ is weakly larger than $\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$ for all menus), but it can also increase the average psychological cost.

Definition 1. There is information avoidance for F if $v(F, \hat{p}) > V(F|\mu)$.

A strict inequality indicates that avoidance must be an "active" choice, hence subject to a strictly positive cost (see Golman et al., 2017).

Information avoidance for F is equivalent to the cost of information $(I(\mu, F) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}))$ being strictly greater than the "material value of information." If, for example, $d = d_\kappa$, information avoidance occurs only if there is no identity trade-off under ignorance.[8] If there is a trade-off, the psychological cost would be κ. Information is weakly valuable because it has a positive material value and it may eliminate the identity trade-off for some posteriors. Thus, the expected cost would be smaller than κ.

Figure 2 suggests a sufficient condition for information avoidance. Consider the purple line in the right panel: it is the smallest concave function that is greater than the value function, called the concave envelope of the value function (see Kamenica and Gentzkow, 2011). There is information avoidance in the moral wiggle room because the concave envelope is equal to the value

8. If $\kappa = \infty$ and $f^*_{F,\hat{p}} \neq f_{F,\hat{p}}$, $v(F, \hat{p}) = -\infty$ and it cannot be strictly greater than $V(F|\mu)$. Suppose that κ is finite. If $f^*_{F,\hat{p}} \neq f_{F,\hat{p}}$, then $v(F, \hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \kappa$. Since $W(\mu, F) \geq \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$ and $\kappa \geq \int_{\Delta\Omega} d_\kappa(f^*_{F,q}, f_{F,q}, q)d\mu(q)$ (because d_κ is either equal to κ or to 0), it follows that $v(F, \hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \kappa \leq W(\mu, F) - \int_{\Delta\Omega} d_\kappa(f^*_{F,q}, f_{F,q}, q)d\mu(q) = V(F|\mu)$. Thus, information is weakly valuable. The online Appendix C contains additional results on information avoidance under parametric restrictions to the cost function.

function at the prior, but it is sufficiently concave when calculated at the posteriors (the purple dot in Figure 2 is $V(a \cup b|\mu)$). This intuition generalizes. Given a menu F, I denote by $\operatorname{cav} v(F, \cdot)$ the concave envelope of $v(F, \cdot)$.

Proposition 1 (Information Avoidance). *If $v(F, \hat{p}) = \operatorname{cav} v(F, \hat{p})$ and the restriction of $\operatorname{cav} v(F, \cdot)$ to the posteriors is not affine,*[9] *then there is information avoidance for F. If there is information avoidance for F, then $d(f^*_{F,q}, f_{F,q}, q) > d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$ for at least one posterior belief q.*

Acquiring information generates a lottery over epistemic situations, so strict "risk aversion" (strict concavity of $v(F, \cdot)$) would imply information avoidance. When the value function is not concave, as in the right panel of Figure 2, one can restrict attention to its concave envelope. The second part of Proposition 1 shows that information aversion requires information to strictly exacerbate the identity trade-off for some posterior beliefs. If, for example, there is an absolute prescription f in F, the inequality in Proposition 1 means information changes at least one payoff-maximizing action (i.e., $d(f^*_{F,q}, f, q) > d(f^*_{F,\hat{p}}, f, \hat{p})$ implies $f^*_{F,q} \neq f^*_{F,\hat{p}}$ for at least one posterior q).

Proposition 1 applies to domains other than social dilemmas. For instance, to explain the well-established evidence that men are less likely than women to have routine medical tests (Courtenay, 2000; Mahalik et al., 2007). Traditional gender identity perpetrates the image that *real men* are "independent, self-reliant, strong, robust and tough" (Courtenay, 2000). A diagnosis of high blood pressure, for example, implies a recommendation (the payoff-maximizing action) to consume less red meat or take leave from work, generating an identity trade-off. Motivated by health-information avoidance, in Section B.1 of Appendix B, I characterize a test that an identity-concerned individual would always take.

9. This condition means that $\operatorname{cav} v(F, \hat{p}) \neq \int_{\Delta\Omega} \operatorname{cav} v(F, q) d\mu(q)$.

3.2. Non-monotonicity of the cost of information: an illustrative example

The moral wiggle room has two simplifying features: information is perfect, and there is a single payoff-maximizing action for all beliefs. The following example introduces a more complex situation to illustrate additional properties of the Identity model.

EXAMPLE 1 (*Poorly informed altruism*). There are two actions: a donation c or no donation n to a charity of unknown quality (either high ω_h or low ω_l, with ex-ante equal probability). The payoffs (in utils) are in Table 2. An

TABLE 2

	ω_h	ω_l
c	8	0
n	0	4

altruistic identity prescribes donation for all posteriors assigning a probability larger than 1/5 to high quality (thus also under ignorance). Otherwise, the prescription is no-donation. A donation maximizes the individual's material payoffs for any belief assigning a probability of at least 1/3 to ω_h (thus, also under ignorance). Therefore, the identity trade-off emerges for any posterior that assigns a probability smaller than 1/3 and larger than 1/5 to ω_h (see Figure A.1 in Appendix A). The individual can acquire information μ that leads to two equally probable posteriors q', q'', with $q'(\omega_h) = 3/4$ and $q''(\omega_h) = 1/4$. Information is costly because the posterior q'' falls into the trade-off region with a probability of 1/2. The value of $c \cup n$ under ignorance is $v(c \cup n, \hat{p}) = 1/2 \cdot 8 - d(c, c, \hat{p}) = 4$ and the value of $c \cup n$ with information is

$$V(c \cup n|\mu) = \frac{1}{2} \cdot [6 - d(c, c, q')] + \frac{1}{2}[3 - d(n, c, q'')] = 4.5 - \frac{1}{2} \cdot d(n, c, q'').$$

So there is information avoidance if $d(n, c, q'') > 1$. Suppose that the individual acquires perfect information $\bar{\mu}$, corresponding to $\bar{\mu}(\delta_{\omega_h}) = \bar{\mu}(\delta_{\omega_l}) = 1/2$. Then,

$$V(c \cup n|\bar{\mu}) = \frac{1}{2} \cdot [8 - d(c, c, \delta_{\omega_h})] + \frac{1}{2} \cdot [4 - d(n, n, \delta_{\omega_l})] = 6,$$

which is strictly larger than $v(c \cup n, \hat{p})$. Thus, perfect information is better than ignorance, which is better than partial information ($V(c \cup n|\bar{\mu}) > v(c \cup n, \hat{p}) > V(c \cup n|\mu)$). Figure 3 illustrates the example. ■

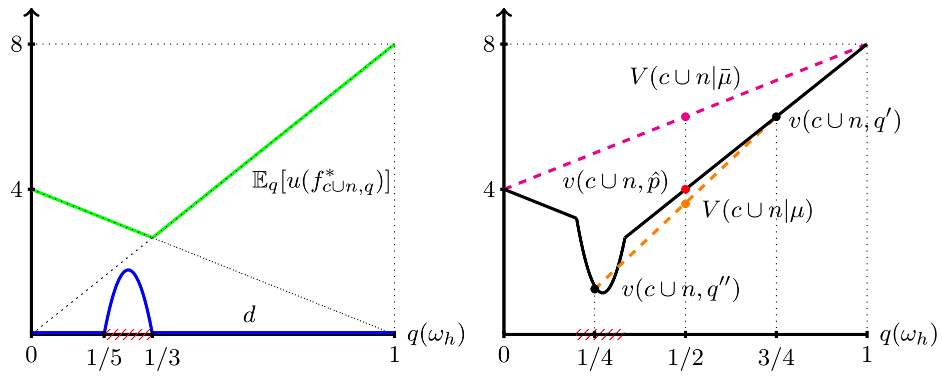

FIGURE 3. Poorly informed altruism. Left panel: the material value of $c \cup n$ as a function of q, $\mathbb{E}_q[u(f^*_{c\cup n,q})] = \max\{\mathbb{E}_q[c], \mathbb{E}_q[n]\}$ (green line), and the psychological cost (blue line). Right panel: the function $v(c \cup n, \cdot)$ (black solid line), the value of $v(c \cup n, \hat{p})$ (red dot), the value of $V(c \cup n|\mu)$ (orange dot), which is the 1/2-1/2 average of $v(c \cup n, q')$ and $v(c \cup n, q'')$. The value of $V(c \cup n|\bar{\mu})$ (purple dot), which is the 1/2-1/2 average of $v(c \cup n, \delta_{\omega_h}) = 8$ and $v(c \cup n, \delta_{\omega_l}) = 4$.

Example 1 highlights the potential non-monotonicity of the cost of information with respect to the quality of information. Typically, better information is more costly (e.g., Sims, 2003; Pomatto et al., 2023), but in the example, μ can be more costly than perfect information (if $d(n, c, q'') > 0$, then $I(\mu, c \cup n) = 1/2d(n, c, q'') > 0 = I(\bar{\mu}, c \cup n)$). Formally, I say that better information is more costly for F if $I(\nu, F) \geq I(\mu, F)$, when an experiment ν consistent with the prior is (Blackwell) more informative than μ (see Definition A.1 in Appendix A).

PROPOSITION 2 (Sufficient and necessary conditions for monotonicity). *If $q \mapsto d(f^*_{F,q}, f_{F,q}, q)$ is convex and continuous, better information is more costly for F. Assume that $\hat{p}$ has full support and $I(\nu, F)$ is finite for all experiments ν consistent with the prior. If better information is more costly for F, then $d(f^*_{F,q}, f_{F,q}, q)$ is convex in q.*

The cost of information can be non-monotone, because better information can eliminate the identity trade-off (as in Example 1). Non-monotonicity complicates the problem of optimal information disclosure to an identity-concerned recipient. Consider a charity that wishes to disclose information about its beneficiaries to potential donors. Providing too much information can have a negative effect and discourage donations, while providing incomplete information may lead donors to "close their eyes" and donate (this is consistent with experimental evidence, see Fong and Oberholzer-Gee, 2011). In Example 1, information avoidance leads to making a donation, while acquiring perfect information leads to donating only half of the time.

Non-monotonicity introduces an asymmetry to the interpretation of information choices from the point of view of an external observer. The rejection of inconvenient information suggests that identity concerns play a role. Conversely, the acquisition of information is inconclusive about the relevance of identity, because worse information could be rejected. This asymmetry is relevant because field and laboratory data are typically one-shot decisions about information acquisition and thus *underestimate* identity concerns. I address this issue in Section 5.1, where I show that observing information preferences in the absence of material gain partially reveals prescriptions.

A second conclusion, derived from Example 1, is that information avoidance may appear *unmotivated*. A pure altruist (who cares about the effectiveness of their donations) may reject instrumental information due to a "modest" identity trade-off. Consistent with this result, Niehaus (2014) found that only 3 percent of donors acquire information prior to donating.[10] In the Identity model, information avoidance is always motivated, because it necessarily requires the existence of an identity trade-off, however "small" it may be. But

10. Similarly, in the context of cooperation, Hoffman et al. (2015) observed that willful ignorance of the cost of cooperation leads to higher cooperation rates. In laboratory experiments, Kandul and Ritov (2017) found that some dictators prefer not to know their payoff and act altruistically, and Andersson et al. (2022) found that when subjects avoid learning a donation norm, they donate more (on average) than non-avoiders.

the presence of a trade-off is a rather common situation, as the only exception is when the payoff-maximizing actions and the prescriptions coincide for *all* posterior beliefs.

4. Avoiding the situation and applications

4.1. Avoiding the situation

To reduce the identity trade-off, the individual has an alternative to willful ignorance: modifying future opportunities. Having more opportunities has a positive material value ($W(\mu, F \cup G)$ is larger than $W(\mu, F)$ for all menus F, G), but can also exacerbate the identity trade-off. If the latter effect is stronger than the former, commitment is optimal. For example, in the moral wiggle room, a reluctant altruist would prefer commitment to a rather than facing $a \cup b$.

Suppose that the prescriptions in a menu F are identical to the prescriptions in $F \cup G$ (i.e., $f_{F,q} = f_{F\cup G,q}$ for all the posteriors and the prior). In this case, I say that $F \cup G$ is *prescriptively equivalent* to F. For example, if f is an absolute prescription in $f \cup g$, then $f \cup g$ is prescriptively equivalent to f. A natural requirement is that the identity trade-off is weakly more costly in $F \cup G$ when it is prescriptively equivalent to F. Adding actions that are not prescriptions (those in G) can only exacerbate the identity trade-off, as these actions can be payoff-maximizing for some beliefs. In this case, namely if $d(f^*_{F\cup G,q}, f, q) \geq d(f^*_{F,q}, f, q)$ for all posteriors and the prior when $F \cup G$ is prescriptively equivalent to F, I say that d is *regular*. Both d_κ and d_e are regular (see Fact A.1 in Appendix A).

Proposition 3 (Avoiding the situation). *Suppose that $F \cup G$ is prescriptively equivalent to F and d is regular. Commitment to F is optimal whenever the additional psychological cost for a posterior q (i.e., $\mu(q)(d(f^*_{F\cup G,q}, f, q) - d(f^*_{F,q}, f, q))$) is larger than the material value of flexibility $W(\mu, F \cup G) - W(\mu, F)$. If commitment to F is strictly optimal, then $f^*_{F\cup G,q} \neq f^*_{F,q}$ for at least one posterior belief q.*

Regularity implies that, even for a single epistemic situation, a sufficiently costly identity trade-off triggers a preference for commitment. Consider d_κ. Commitment to F is strictly valuable only if there is no identity trade-off in F, but there is in $F \cup G$. Thus, G must contain at least one action that is optimal for a posterior, but this action is not the prescription. A less extreme case is $d = d_e$, where the result depends on the slope of φ (see Corollary C.2 in Appendix C).

Regularity of d combined with prescriptive equivalence between $F \cup G$ and F ensures monotonicity of the psychological cost with respect to flexibility (i.e., $I(\mu, F \cup G) \geq I(\mu, F)$). When d is not regular or the menus are not prescriptively equivalent, this monotonicity may fail. As with information, non-monotonicity introduces an asymmetry to the interpretation of situation choices. Observing commitment may signal identity concerns, whereas a preference for flexible situations is inconclusive about the relevance of identity. Thus, laboratory and field data about one-shot avoidance of situations (e.g., DellaVigna et al., 2012; Andreoni et al., 2017; Schwartz et al., 2021) *underestimate* identity concerns. In Section 5.2, I show what types of situation choices are useful to infer prescriptions.

4.2. Avoidance of situations: applications

Excess entry into competitive environments (even without overconfidence). An individual has to decide whether to enter e a competitive environment or not n. Uncertainty concerns the returns of entering, while the value of not entering is 0. The individual can commit to entering or maintain flexibility $N = e \cup n$. The Identity model is consistent with a preference for commitment $V(e|\mu) > V(N|\mu)$. More importantly, this preference does not require *overconfidence*, because it can occur even if the expected utility of entering is lower than that of not entering (i.e., $\mathbb{E}_{\hat{p}}[u(e)] \leq u(n) = 0$).[11]

11. Suppose that the identity absolutely prescribes entering the environment. Then $V(N|\mu) = \int_{\Delta\Omega} \max\{\mathbb{E}_q[u(e)], 0\}\, d\mu(q) - \int_{\Delta\Omega} d(f^*_{N,q}, e, q) d\mu(q)$. Thus, $V(e|\mu) > V(N|\mu)$

Therefore, the Identity model rationalizes (1) excess entry into new markets. An entrepreneur can strictly prefer to enter a new market because of a desire to protect their identity of being bold (Brocas and Carrillo, 2004), even if they know that the investment has a negative Net Present Value ($\mathbb{E}_{\hat{p}}[u(e)] \leq 0$); and (2) gender-driven sorting into competitive environments, which is often ascribed to the overconfidence of men (Niederle and Vesterlund, 2007). However, some males enter the competitive environment to avoid the identity trade-off, even if they know they would be better off doing otherwise.[12]

Opting out of social dilemmas. Consider the payoffs of Example 1. The identity model is consistent with $V(c|\mu) > V(n|\mu) > V(c \cup n|\mu)$, which holds if $5 < d(n, c, q')$. Ideally, the individual would commit to making a donation, but if this is unfeasible, committing to not donating (for example by avoiding a fundraiser) may be better than having flexibility. In the field experiments of DellaVigna et al. (2012) and Andreoni et al. (2017), potential donors are unable to commit to making a donation. Therefore, there may be individuals among those who avoid meeting the fundraiser who would have been better off if given the opportunity to commit to making a donation. A preference for commitment applies also to situations without uncertainty. In that case, information avoidance is inapplicable, and commitment is the only possible strategy to reduce the identity trade-off. Therefore, the Identity model can explain the results of Dana et al. (2006), where some dictators sacrifice a monetary amount to avoid entering a standard dictator game (see Section B.3 of Appendix B).

Gender-related preference for flexibility. Consider the so-called *flexibility stigma*: for a male worker, asking for work flexibility, especially for family

---

whenever $\int_{\Delta\Omega} d(f^*_{N,q}, e, q) d\mu(q) > \int_{\Delta\Omega} \max\{\mathbb{E}_q[u(e)], 0\} \, d\mu(q) - \mathbb{E}_{\hat{p}}[u(e)]$. The last inequality can be satisfied even if $\mathbb{E}_{\hat{p}}[u(e)]$ is negative.

12. van Veldhuizen (2022) argues that the gender gap in competitiveness derives from differences in risk aversion and self-confidence rather than competitiveness traits. Identity concerns can still play a role if being male prescribes being a "risk lover" or "self-confident".

caregiving, is an impermissible lack of commitment, if not a feminine behavior (see Williams et al. 2013, Rudman and Mescher 2013, and Vandello et al. 2013). If there is uncertainty about the value of doing childcare, learning that doing it is better than delegating it generates the identity trade-off (for a traditional masculine identity). To reduce the identity trade-off, the worker avoids flexibility.

5. Inferring prescriptions from choices

Identity is typically unobservable from an external point of view, as are the prescriptions.[13] Although in many applied situations, such as the moral wiggle room, the prescriptions are rather transparent, in more complex situations they are not. Moreover, the non-monotonicities described in the previous sections further complicate the task of identifying prescriptions from choices. In this section, I provide two possible solutions: the first exploits information choices; the second, preferences toward situations. Apart from providing new tools for experimental works on identity, the two results help to complete the analogy between information and opportunity choices.

5.1. Inference from information choices and the "demand for beliefs"

The intuition motivating the first approach comes from the moral wiggle room. Suppose that an action in a menu F delivers higher material utility than the other available actions in all states (as a does in the moral wiggle room example assuming $u(x, y) = x$). I call this action "payoff-dominant" in F.[14]

13. Few papers explicitly attempt to infer identity from behavior: Krupka and Weber (2013) use coordination games to identify social norms, Atkin et al. (2021) propose an empirical approach based on revealed food choice, and Ballester and Bozbay (2021) provide a theoretical revealed preference analysis of social identity. Piermont (2019) axiomatizes a model in which the individual cares about the signal that a choice conveys in terms of its possible rationalizations.

14. An action $f^* \in F$ is payoff-dominant in F if $u(f^*(\omega)) \geq u(g(\omega))$ for all $\omega \in \Omega$ and all $g \in F$. It follows that $f^*_{F,q} = f^*$ for all beliefs, because $\mathbb{E}_q[u(f^*)] \geq \mathbb{E}_q[u(g)]$ for all $q \in \Delta\Omega$ and all

With a payoff-dominant action, information has no material value because the payoff-maximizing action is independent of beliefs. Therefore, observing willful ignorance implies that an alternative action must generate the identity trade-off for at least one posterior. Observing information acquisition implies that the payoff-dominant action cannot be the prescription under ignorance, otherwise ignorance would be optimal.

Proposition 4 (Inferring prescriptions from information choices). *Assume that f^* is payoff-dominant in $f \cup f^*$. Information avoidance for $f \cup f^*$ implies that f is the q-belief prescription in $f \cup f^*$ for at least one posterior belief q. If information is strictly valuable for $f \cup f^*$, then f is the $\hat{p}$-belief prescription in $f \cup f^*$.*

Proposition 4 shows that information acquisition (or avoidance) in the Identity model can respond to a “demand for belief” (see Loewenstein and Molnar, 2018). Even if it has no instrumental value, information is costly (or valuable) because it changes the prescriptions, thus affecting the identity trade-off. A more subtle case of demand for beliefs derives from the properties of the cost function. Information that changes neither the payoff-maximizing action *nor the prescription* can still be costly (or valuable) because the psychological cost varies with the posteriors.

Proposition 5 (Belief-dependent utility). *If all the posteriors belong to a trade-off region of F and $d(f, f', q)$ is strictly convex (resp. concave) in q in that region, then information is strictly costly (resp. valuable).*

The condition of Proposition 5 holds for d_e when φ is strictly convex, but not for d_κ, for example.

$g \in F$. Moreover, information has no material value because $W(\mu, F) = \int_{\Delta\Omega} \mathbb{E}_q[u(f^*)]d\mu(q) = \mathbb{E}_{\hat{p}}[u(f^*)] = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$.

5.2. *Inference from choices of opportunities and the demand for non-instrumental flexibility*

Preferences over future opportunities are also informative about prescriptions. In the moral wiggle room, a dictator who prefers committing to the action a, rather than facing the more flexible situation $a \cup b$, reveals that b generates a trade-off for some posterior beliefs. In order to infer prescriptions in menus with more than two actions, however, an additional property is required. I say that the prescriptions are *context-independent*, if adding an action g to a menu F in which f is the q-belief prescription, implies that either f is the q-belief prescription in $F \cup g$ or g becomes the q-belief prescription in $F \cup g$. Context-independence rules out the case, for example, in which adding an action c in the moral wiggle room, turns a into the prescription in state ω_2. If prescriptions are context-independent, either c becomes the prescription or the prescription is b. Context-independence trivially holds when F is a singleton.

Proposition 6 (Inferring prescriptions from choices of opportunities). *Assume that $f^* \neq g$, f^* is payoff dominant in $F \cup g$ and the prescriptions are context-independent. If $v(F, \hat{p}) \neq v(F \cup g, \hat{p})$, then g is the $\hat{p}$-belief prescription in $F \cup g$. If $V(F|\mu) \neq V(F \cup g|\mu)$, then g is the q-belief prescription in $F \cup g$ for at least one posterior belief q.*

The presence of a payoff-dominant action f^* in $F \cup g$ equalizes the material values of F and $F \cup g$, so any difference in their valuations must come from the identity trade-off. Context-independence ensures that any variation in the identity trade-off is due to g. Proposition 6 shows that choices over future opportunities are as informative about identity as willful ignorance is. Therefore, they can be a key component in the design of experiments on identity, either alone or in conjunction with choices about information.

I conclude this section by completing the analogy between information and flexibility choices. The identity trade-off varies as either or both the prescription and the payoff-maximizing action change. Among the possible combinations, there are two extreme cases: (1) the prescriptions are fixed and the payoff-maximizing actions vary, and (2) the payoff-maximizing actions are fixed and the prescriptions vary. The demand for commitment in Proposition 3 is analogous to information avoidance in the presence of an absolute prescription, and both are examples of (1). In these cases, information and flexibility do not change the prescriptions, but can affect the identity trade-off by changing the payoff-maximizing actions. Proposition 6 is analogous to a "demand for beliefs," and both are examples of (2). Indeed, Proposition 6 displays a preference for non-instrumental flexibility (or commitment): a desire for including (or excluding) an action in a menu, even if the action is never payoff-maximizing.[15] In both cases, information and flexibility do not change the payoff-maximizing actions, but can affect the identity trade-off by changing the prescriptions.

6. Extension: Meta-prescriptions

Prescriptions are often more general than actions. In a moral dilemma with resolvable uncertainty, an altruist identity can establish that the appropriate behavior is *learning* (e.g., Bartling et al., 2014; Grossman and van der Weele, 2016). Certain religions sanction the mere possibility of acquiring information (e.g., by possessing certain books), rather than its use. Furthermore, prescriptions can be about future opportunities, such as, for a married woman, not entering the labor market (e.g., Bertrand et al., 2015), or for a black student, not accumulating human capital (e.g., Austen-Smith and Fryer Jr, 2005) (see Section C.3 in the online Appendix). In this section, I outline an extension of the Identity model that captures a general notion of prescription. Consider a menu F. The Identity model assigns values to

15. For an example of non-instrumental flexibility, suppose that there are only two possible tipping options, 2% and 15%. A reluctant tipper maximizes their payoff by tipping 2%, but the prescription is 15%. Adding a third option, say 7%, may be strictly valuable if it becomes the prescription, even if tipping 2% is still payoff-maximizing. An example of non-instrumental commitment is a preference for a over $a \cup b$ in the moral wiggle room.

F with information acquisition, denoted by F_μ, and to F under ignorance, denoted by $F_{\delta_{\hat{p}}}$. To model meta-prescriptions, I add an ex-ante stage, where the individual evaluates the generalized "menus" $\mathbb{F} = \{F_\mu, F_{\delta_{\hat{p}}}\}$ and $\{F_{\delta_{\hat{p}}}\}$. A meta-prescription is an element of $\mathbb{F}$. For example, the meta-prescription of acquiring information in $\mathbb{F}$ is F_μ. The interpretation is that, by choosing $\mathbb{F}$, the individual does not exclude the possibility of acquiring information at a later stage. On the contrary, committing to $F_{\delta_{\hat{p}}}$ excludes this possibility. Notice that choosing $\mathbb{F}$ does not imply the acquisition of information, but simply having *the possibility* of acquiring it. A simple functional form that extends the Identity model to include meta-prescriptions is the following:

$$\hat{V}(\mathbb{F}) = \max\{V(F|\mu), v(F, \hat{p})\} - D(H^*_{\mathbb{F}}, H_{\mathbb{F}}), \tag{3}$$

where D is a positive function with $D(H, H) = 0$. The expression $H^*_{\mathbb{F}}$ denotes the pair that maximizes the second-stage utility of the individual (either F_μ or $F_{\delta_{\hat{p}}}$), and $H_{\mathbb{F}}$ the meta-prescription in $\mathbb{F}$. The Identity model corresponds to $D = 0$. In the generalized model, the identity trade-off emerges also at the ex-ante stage.

The meta-prescription of acquiring information implies $D(H^*_{\mathbb{F}}, F_\mu)$ in Equation (3). There are two possible cases: $V(F|\mu) \geq v(F, \hat{p})$ and $v(F, \hat{p}) \geq V(F|\mu)$. In the first case, $\hat{V}(\mathbb{F}) = \hat{V}(F_\mu) \geq \hat{V}(F_{\delta_{\hat{p}}})$. Anticipating information acquisition if given the possibility, the individual dislikes committing to ignorance, and is indifferent between flexibility and commitment to F_μ. In the second case, $\hat{V}(\mathbb{F}) = v(F, \hat{p}) - D(F_{\delta_{\hat{p}}}, F_\mu)$, which implies $\hat{V}(\mathbb{F}) \leq \hat{V}(F_{\delta_{\hat{p}}})$. Anticipating a future violation of the meta-prescription, the individual prefers to commit to ignorance rather than having the possibility of choosing later. A reluctant altruist entering a moral wiggle room may prefer to commit to a situation where learning the state before playing is impossible, even if they know that they will not learn it if given the possibility. Note that a strict preference for commitment to $F_{\delta_{\hat{p}}}$ implies $v(F,\hat{p}) > V(F|\mu)$, thus a "second-stage" identity concern, of the type formalized in the Identity model, must play a role.[16] Similar considerations apply to a meta-prescription of ignorance.

7. Related literature

The seminal works of Akerlof and Kranton (2000, 2010) introduced identity in economics.[17] In the model of Akerlof and Kranton (2000), the utility of an individual is a function of their own actions, the actions of others, and their own identity. More specific models of identity belong to two distinct, but not mutually exclusive, approaches.

The first is based on identity uncertainty. Köszegi (2006) introduces a model where the individual's utility depends on their beliefs about their ability, which can be self-signalled by taking ambitious actions. Bénabou and Tirole (2011) study moral behavior where individuals are unsure about their "deep values," such as being altruistic, and take (costly) actions to self-signal identity. The model of Bénabou and Tirole (2011) predicts a specific form of information avoidance with regard to the price of "taboo" transactions. Grossman and van der Weele (2016) take a dual-self approach in which the observer-self does not know the type of the doer-self. Acting altruistically has uncertain benefits, but the individual can acquire perfect information. In one equilibrium, the altruistic types prefer willful ignorance, so as to avoid pooling with "low" types. Their model features multiple equilibria and the willful ignorance equilibrium requires special parametric restrictions. All these models consider simplified settings (e.g., two-action, two-state), while the Identity model accounts for menus of actions, general uncertainty, and general information. The latter

16. Indeed, if the psychological cost in the Identity model is zero, $V(F|\mu) = W(\mu, F)$ and $v(F,\hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$, which contradicts $v(F,\hat{p}) > V(F|\mu)$.

17. Applications include organization theory (Akerlof and Kranton, 2005), political economy (Bonomi et al., 2021), finance (D'Acunto, 2019), labor (Bertrand et al., 2015; Oh, 2021), preferences estimation (Benjamin et al., 2010), and consumption (Atkin et al., 2021).

feature is central in order to capture the non-monotonicity of the psychological cost, for example. Moreover, the Identity model does not make assumptions about the nature of the identity trade-off, meaning its predictions hold even in the absence of uncertainty about identity and when actions do not signal values.

The second approach is based on cognitive dissonance (Festinger, 1962). Konow (2000) studies two-person allocation decisions under certainty, and Nyborg (2011), Matthey and Regner (2011), Spiekermann and Weiss (2016), Ellingsen and Mohlin (2019), and Momsen and Ohndorf (2020) use models of cognitive dissonance to rationalize experimental evidence in various domains. Gilboa et al. (2022) axiomatize a deterministic model of consumption in which the presence of "prohibited" substances, such as meat for a vegetarian, is costly. These models focus on the value of actions, whereas the Identity model jointly treats information and situation choices.

The paper contributes to the literature on information aversion (see the reviews of Hertwig and Engel 2016, Golman et al. 2017, and Sunstein 2020). Among the many rationalizations of information aversion,[18] the Identity model retains both expected utility and Bayesian updating, as in rational inattention (e.g., Sims, 2003; Pennesi, 2015; De Oliveira et al., 2017). Rational inattention, however, is inconsistent with a preference for commitment.

The paper also contributes to the literature on menu choice in the presence of information acquisition or identity concerns. Epstein (2006) and Epstein et al. (2008) develop models of sophisticated non-Bayesian updating that predict a desire for commitment. In the Identity model, the preference for commitment derives from a desire to reduce the identity trade-off. In Dillenberger and Sadowski (2012), a preference for commitment derives from aversion to shame, but their model does not consider uncertainty or information. In Section B.3 of Appendix B, I apply the Identity model to

18. Golman et al. (2017) list 13 non-strategic motivations for information avoidance. See also Trimmer et al. (2020) for a review of models predicting willful ignorance.

explain the evidence motivating Dillenberger and Sadowski (2012). Lastly, Alaoui (2016) proves that the preservation of self-image may lead to a preference for smaller menus (of lotteries), in a model where lotteries convey information about the unobservable ability of the decision-maker.

8. Concluding remarks

Given the focus of my model, I have left some aspects for future research: first, a natural extension would include costly "selection" of identity (as in Rabin, 1995). Identity is malleable, and an individual can partly select their preferred identity in a choice situation. However, it is not clear how identity selection will interact with information and/or the flexibility of a situation. In particular, the timing of identity selection is crucial, as it can be selected before or after information arrives (and before or after the choice of future flexibility). A second aspect worth exploring is strategic interactions. In the original model of Akerlof and Kranton (2000), the actions of others enter the utility through their material consequences and identity considerations. The Identity model is applicable to strategic interactions, bearing in mind that identity concerns affect the value of situations, while the value of an action depends on its expected utility.

Appendix A: Proofs

A.1. A bound to the number of trade-off regions

Given a menu F, for each action f in F, I denote by P_f^F the possibly empty convex set of beliefs for which f is the prescription, i.e., $P_f^F = \{q \in \Delta\Omega : f_{F,q} = f\}$. Similarly, each action f in F determines a possibly empty convex set B_f^F containing all beliefs that make it the payoff-maximizing action, i.e., $B_f^F = \left\{q \in \Delta\Omega : f^*_{F,q} = f\right\}$. The identity trade-off emerges in the epistemic situation (F, q) if $q \in P_f^F \cap B_g^F$ for some $f \neq g$. Therefore, I define the trade-off regions for F as the non-empty (and convex) intersections $P_f^F \cap B_g^F$ for some $f \neq g$. For example, in the moral wiggle room of Figure 1, $P_a^{a\cup b} = \{q \in \Delta\Omega : q(\omega_2) \leq 0.75\}$, $P_b^{a\cup b} = \{q \in \Delta\Omega : q(\omega_2) > 0.75\}$, $B_a^{a\cup b} = \Delta\Omega$ and $B_b^{a\cup b} = \emptyset$. The trade-off region is $P_b^{a\cup b} \cap B_a^{a\cup b} = P_b^{a\cup b} \cap \Delta\Omega = P_b^{a\cup b}$. Let P^F denote the family of non-empty sets P_f^F for $f \in F$ and let B^F be the family of non-empty sets B_f^F for $f \in F$. Lastly, I define the set $N^F = \left\{f \in F : P_f^F \in P^F \text{ and } B_f^F \in B^F\right\}$, which contains the actions that are both payoff-maximizing for some posterior beliefs and prescription for some other (possibly different) posterior beliefs. In the moral wiggle room $N^{a\cup b} = \{a\}$.

Proposition A.1 (Maximum number of trade-off regions). *Given a menu F, there are at most $|P^F| \cdot |B^F| - |N^F|$ trade-off regions for F.*

Proof of Proposition A.1. Consider the two families P^F and B^F of non-empty subsets of $\Delta\Omega$. Each set $P_f^F \in P^F$ has at most $|B^F|$ non-empty intersections with the sets in B^F. Therefore, there are at most $|P^F| \cdot |B^F|$ non-empty intersections. However, if for some $f \in F$, $P_f^F \in P^F$ and $B_f^F \in B^F$, the intersection $P_f^F \cap B_f^F$ is not a trade-off region. Therefore, there are at most $|P^F| \cdot |B^F| - |N^F|$ trade-off regions for F. □

Suppose that an external observer has no information about prescriptions and payoff-maximizing actions. In this case, all actions can potentially be prescriptions and payoff-maximizing for some beliefs. Therefore, if $|F| = n$, then $|P^F| = |B^F| = |N^F| = n$ and Proposition A.1 implies that the maximum number of trade-off region is $n(n-1)$. Suppose that there is a unique payoff-maximizing action in a menu with n actions. Then $|B^F| = 1$, $|P^F| = n$, and the upper bound becomes $n - 1$ because N^F contains at most the payoff-maximizing action. Symmetrically, the upper bound $n - 1$ holds if there is an absolute prescription (i.e., $|P^F| = 1$).

A.2. Proofs of the results in the main text

Before proving Proposition 1, I define the concave envelope of the value function:

$$\operatorname{cav} v(F, q) = \inf \left\{ h(q) : h \text{ is affine, continuous and } h(q') \geq v(F, q'),\ \forall q' \right\}.$$

Proof of Proposition 1. For the first part, suppose that $v(F, \hat{p}) = \operatorname{cav} v(F, \hat{p})$ and the restriction of $\operatorname{cav} v(F, \cdot)$ to the posteriors is not affine. By the concavity of $\operatorname{cav} v(F, \cdot)$ and the monotonicity of the integral $v(F, \hat{p}) = \operatorname{cav} v(F, \hat{p}) \geq \int_{\Delta\Omega} \operatorname{cav} v(F, q) d\mu(q) \geq \int_{\Delta\Omega} v(F, q) d\mu(q) = V(F|\mu)$. Thus, there is either information avoidance for F (if the inequality is strict) or the information is irrelevant (if it is an equality). However, in the case of equality, it means that the restriction of $\operatorname{cav} v(F, \cdot)$ to $\operatorname{supp} \mu$ is affine, contradicting the hypothesis. For the second part, if there is information avoidance for F, then $I(\mu, F) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$ is strictly larger than $W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] \geq 0$. Therefore, $\int_{\Delta\Omega} d(f^*_{F,q}, f_{F,q}, q) d\mu(q) > d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$, implying that, for at least one posterior belief q, $d(f^*_{F,q}, f_{F,q}, q) > d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$. □

Trade-off regions in Example 1.

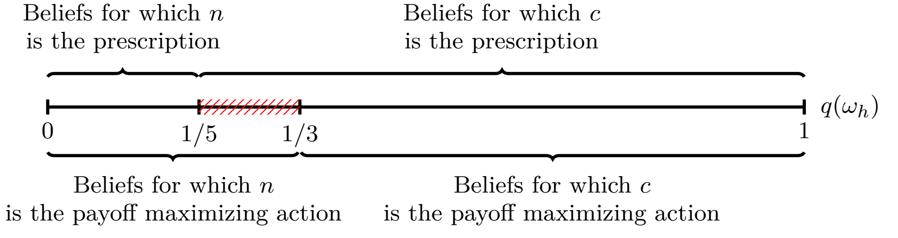

FIGURE A.1. The black solid line is the probability of state ω_h. The prescription is n for each belief that assigns a probability of less than 1/5 to state ω_h (i.e., $f_{c \cup n, q} = n$ for all q such that $q(\omega_h) \leq 1/5$), otherwise the prescription is c. The action n maximizes the material payoff for each belief that assigns a probability of less than 1/3 to state ω_h, otherwise the payoff-maximizing action is c. The red pattern highlights the trade-off region.

Before proving Proposition 2, I introduce the Blackwell (1953) comparative definition of informativeness. Let $\Gamma(\hat{p})$ denote the family of all Bayesian experiments consistent with a prior $\hat{p} \in \Delta\Omega$.

DEFINITION A.1. Given $\mu, \nu \in \Gamma(\hat{p})$, ν is Blackwell more informative than μ, written $\nu \trianglerighteq \mu$, if

$$\int_{\Delta\Omega} \varphi(q) d\nu(q) \geq \int_{\Delta\Omega} \varphi(q) d\mu(q)$$

for all convex and continuous functions $\varphi : \Delta\Omega \to \mathbb{R}$.

Proof of Proposition 2. For simplicity, I denote by $d_F(q)$ the function $d(f^*_{F,q}, f_{F,q}, q)$. For the sufficiency part, by Definition A.1, if $d_F(q)$ is convex and continuous $I(\nu, F) = \int_{\Delta\Omega} d_F(q)d\nu(q) \geq \int_{\Delta\Omega} d_F(q)d\mu(q) = I(\mu, F)$ if $\nu \trianglerighteq \mu$. Thus, better information is more costly for F. For the necessary part, I build on the proof of Lemma 3 in Lipnowski and Mathevet (2018). Suppose that $d_F(q)$ is not convex, then there are $p, p' \in \Delta\Omega$ and $\gamma \in (0, 1)$ such that $d_F(\gamma p + (1-\gamma)p') < \gamma d_F(p) + (1-\gamma)d_F(p')$. Moreover, by setting $\varepsilon = \min_{\omega\in\Omega} \hat{p}(\omega)$ (which is strictly larger than zero by the full support assumption), it holds that $\varepsilon(\gamma p + (1-\gamma)p') \leq \hat{p}$. Now, I define $q = \frac{1}{1-\varepsilon}(\hat{p} - \varepsilon(\gamma p + (1-\gamma)p'))$, $\nu = (1-\varepsilon)\delta_q + \varepsilon(1-\gamma)\delta_p + \varepsilon\gamma\delta_{p'}$, and $\nu' = (1-\varepsilon)\delta_q + \varepsilon\delta_{\gamma p+(1-\gamma)p'}$. By construction, $\nu, \nu' \in \Gamma(\hat{p})$ and $\nu \trianglerighteq \nu'$, but $I(\nu, F) - I(\nu', F) = \varepsilon[(1-\gamma)d_F(p) + \gamma d_F(p') - d_F(\gamma p + (1-\gamma)p')] < 0$, a contradiction to the assumption that better information is more costly for F. □

Fact A.1. *The cost functions d_κ and d_e are regular.*

Proof of Fact A.1. Consider d_e first. Notice that, for any q in the support of μ or $q = \hat{p}$, $\mathbb{E}_q[u(f^*_{F\cup G,q})] = \max_{f\in F\cup G} \mathbb{E}_q[u(f)] \geq \max_{f\in F} \mathbb{E}_q[u(f)] = \mathbb{E}_q[u(f^*_{F,q})]$. Since φ is increasing and $F \cup G$ is prescriptively equivalent to F, it holds that

$$\begin{aligned} d_e(f^*_{F\cup G}, f, q) =& \varphi\left(\max_{f\in F\cup G} \mathbb{E}_q[u(f)] - \mathbb{E}_q[u(f)]\right) \\ \geq& \varphi\left(\max_{f\in F} \mathbb{E}_q[u(f)] - \mathbb{E}_q[u(f)]\right) = d_e(f^*_F, f, q). \end{aligned}$$

Consider $d = d_\kappa$. It is sufficient to show that it cannot be the case that, for some $q \in \operatorname{supp}\mu$ or $q = \hat{p}$, $d_\kappa(f^*_{F,q}, f, q) = \kappa$ and $d_\kappa(f^*_{F\cup G,q}, f, q) = 0$. Suppose that $d_\kappa(f^*_{F,q}, f, q) = \kappa$ for some $f \in F$. This means that $f^*_{F,q} = g \neq f$. If $d_\kappa(f^*_{F\cup G,q}, f, q) = 0$, then $f^*_{F\cup G,q} = f$, which implies $f^*_{F\cup G,q} \in F$ because $f \in F$. However, the inequality $f^*_{F\cup G,q} = f \neq f^*_{F,q} = g$ contradicts the fact that g is the payoff maximizing action in F. □

Proof of Proposition 3. Since $F \cup G$ is prescriptively equivalent to F and d is regular, $d(f^*_{F\cup G,q}, f, q) \geq d(f^*_{F,q}, f, q)$ for all the posterior beliefs and the prior. For the first part, if $\mu(q)\left(d(f^*_{F\cup G,q}, f, q) - d(f^*_{F,q}, f, q)\right) \geq (W(\mu, F\cup G) - W(\mu, F))$ for a posterior belief q, then

$$\begin{aligned} I(\mu, F\cup G) - I(\mu, F) =& \int_{\Delta\Omega}\left[d(f^*_{F\cup G,p}, f, p) - d(f^*_{F,p}, f, p)\right] d\mu(p) \\ =& \mu(q)(d(f^*_{F\cup G,q}, f, q) - d(f^*_{F,q}, f, q)) + \sum_{p\in\operatorname{supp}\mu\setminus q}\left[d(f^*_{F\cup G,p}, f, p) - d(f^*_{F,p}, f, p)\right]\mu(p) \\ \geq& W(\mu, F\cup G) - W(\mu, F) + \sum_{p\in\operatorname{supp}\mu\setminus q}\left[d(f^*_{F\cup G,p}, f, p) - d(f^*_{F,p}, f, p)\right]\mu(p). \end{aligned}$$

Since d is regular, the terms in the square brackets are all positive, thus

$$W(\mu, F \cup G) - W(\mu, F) + \sum_{p \in \text{supp } \mu \setminus q} \left[d(f^*_{F\cup G,p}, f, p) - d(f^*_{F,p}, f, p) \right] \mu(p) \geq W(\mu, F \cup G) - W(\mu, F).$$

This implies $I(\mu, F \cup G) - I(\mu, F) \geq W(\mu, F \cup G) - W(\mu, F)$ and rearranging gives the result. For the second part, if $V(F|\mu) > V(F \cup G|\mu)$ then $I(\mu, F \cup G) - I(\mu, F) > W(\mu, F \cup G) - W(\mu, F) \geq 0$. The inequality $I(\mu, F \cup G) - I(\mu, F) > 0$ implies $\int_{\Delta\Omega}(d(f^*_{F\cup G,q}, f_{F\cup G,q}, q) - d(f^*_{F,q}, f_{F,q}, q))d\mu(q) > 0$, thus for at least one $q' \in \operatorname{supp} \mu$, $d(f^*_{F\cup G,q'}, f_{F\cup G,q'}, q') > d(f^*_{F,q'}, f_{F,q'}, q')$. Since $F \cup G$ is prescriptively equivalent to F, $f_{F,q} = f_{F\cup G,q}$ for all $q \in \operatorname{supp} \mu$. Therefore, $d(f^*_{F\cup G,q'}, f, q') > d(f^*_{F,q'}, f, q')$, which implies $f^*_{F\cup G,q'} \neq f^*_{F,q'}$ for at least one posterior belief q'. □

Proof of Proposition 4. For the first part, there are two cases to check: $f^* = f_{f\cup f^*,\hat{p}}$ and $f = f_{f\cup f^*,\hat{p}}$. In the first case, $v(f \cup f^*, \hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*)]$. Therefore, if $\mathbb{E}_{\hat{p}}[u(f^*)] > \mathbb{E}_{\hat{p}}[u(f^*)] - \int_{\Delta\Omega} d(f^*, f_{f\cup f^*,q}, q)d\mu(q)$, it must be that $\int_{\Delta\Omega} d(f^*, f_{f\cup f^*,q}, q)d\mu(q) > 0$. This means that $f = f_{f\cup f^*,q}$ for at least one posterior belief $q \in \operatorname{supp} \mu$. In the second case, $v(f \cup f^*, \hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*)] - d(f^*, f, \hat{p})$. Therefore, if $\mathbb{E}_{\hat{p}}[u(f^*)] - d(f^*, f, \hat{p}) > \mathbb{E}_{\hat{p}}[u(f^*)] - \int_{\Delta\Omega} d(f^*, f_{f\cup f^*,q}, q)d\mu(q)$, it means that $d(f^*, f, \hat{p}) < \int_{\Delta\Omega} d(f^*, f_{f\cup f^*,q}, q)d\mu(q)$. Suppose that, for no $q \in \operatorname{supp} \mu$, $f = f_{f\cup f^*,q}$, then $d(f^*, f, \hat{p}) < \int_{\Delta\Omega} d(f^*, f^*, q)d\mu(q) = 0$, a contradiction. For the second part, if $v(f^* \cup f, \hat{p}) > V(f^* \cup f|\mu)$, then $\int_{\Delta\Omega} d(f^*, f_{f\cup f^*,q}, q)d\mu(q) < d(f^*, f_{f^*\cup f,\hat{p}}, \hat{p})$, which implies $f^* \neq f_{f^*\cup f,\hat{p}}$. □

Proof of Proposition 5. Suppose that all the posteriors belong to $P^F_f \cap B^F_g$, for some $f, g \in F$ with $f \neq g$. The convexity assumption on the prescriptions implies $f = f_{F,\hat{p}}$. Then, strict convexity (or concavity) of d in q and the fact that μ is Bayesian consistent with $\hat{p}$ imply $V(F|\mu) = \mathbb{E}_{\hat{p}}[u(g)] - \int_{\Delta\Omega} d(g, f, q)d\mu(q) < (>) \mathbb{E}_{\hat{p}}[u(g)] - d(g, f, \hat{p}) = v(F, \hat{p})$. □

Proof of Proposition 6. Suppose that $V(F|\mu) = \mathbb{E}_{\hat{p}}[u(f^*)] - I(\mu, F) \neq \mathbb{E}_{\hat{p}}[u(f^*)] - I(\mu, F \cup g) = V(F \cup g|\mu)$. It must be that $I(\mu, F) \neq I(\mu, F \cup g)$, or $\int_{\Delta\Omega} d(f^*, f_{F\cup g,q}, q)d\mu(q) \neq \int_{\Delta\Omega} d(f^*, f_{F,q}, q)d\mu(q)$, which implies $d(f^*, f_{F,q}, q) \neq d(f^*, f_{F\cup g,q}, q)$ for at least one posterior belief q. By context-independence of the prescriptions, it means that $g = f_{F\cup g,q}$ for some $q \in \operatorname{supp} \mu$. □

Appendix B: Additional results

B.1. Disclosure to an identity-caring individual

Motivated by the gender gap in preventive healthcare discussed in Section 3.1, I study the existence and the properties of a test (i.e., a Bayesian experiment) that is always acquired by an identity-concerned individual. For a given menu F, consider a (non-null) *preferred state* of the individual in F. This is a state of the world that, with strictly positive probability according to the prior, gives the best payoff in F. For example, with the payoffs of Table 2, a preferred state is ω_h. Indeed, the highest payoff in $c \cup n$ is 8 and it occurs in state ω_h that has a non-zero probability according to the prior. Formally, a preferred state is $\bar{\omega} \in \Omega$ such that $x = f(\bar{\omega})$ and $u(x) \geq \max_{g \in F} \max_{\omega \in \Omega : \hat{p}(\omega) > 0} u(g(\omega))$. Preferred states can be multiple. A menu is *balanced* if $u(f^*_{F,\hat{p}}(\bar{\omega})) \leq \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$, namely when the payoff-maximizing action under ignorance is not optimal in a preferred state $\bar{\omega}$. I call the cost *ε-flat* at $\hat{p}$ if $d(f, f_{F,\hat{p}}, \hat{p}) = d(f, f_{F,q}, q)$ when $\max_{\omega \in \Omega} |\hat{p}(\omega) - q(\omega)| \leq \varepsilon$. This condition holds, for example, when there is an absolute prescription in F and $d(\cdot, \cdot, p) = d(\cdot, \cdot, q)$ for all $p, q \in \Delta\Omega$. The following result characterizes the optimal test:

Proposition B.1 (Always-acquired test). *If F is balanced and $u(f(\bar{\omega})) - d(f^*_{F,\delta_{\bar{\omega}}}, f_{F,\delta_{\bar{\omega}}}, \delta_{\bar{\omega}}) > v(F, \hat{p})$ for a preferred state $\bar{\omega}$, there is $\alpha > 0$ and a $\mu_\alpha \in \Gamma(\hat{p})$ that is always acquired (i.e. $V(F|\mu_\alpha) > v(F, \hat{p})$) when d is $\frac{\alpha}{1-\alpha}$-flat at $\hat{p}$. The test μ_α is defined as $\mu_\alpha = \alpha\delta_{\delta_{\bar{\omega}}} + (1-\alpha)\delta_{\tilde{p}}$ where $\tilde{p} \in \Delta\Omega$ is*

$$\tilde{p}(\omega) = \begin{cases} \frac{\hat{p}(\omega)}{1-\alpha} & \text{if } \omega \neq \bar{\omega} \\ \frac{\hat{p}(\omega)-\alpha}{1-\alpha} & \text{if } \omega = \bar{\omega}. \end{cases}$$

Proof of Proposition B.1. Consider $\mu_\alpha \in \Delta\Delta\Omega$ defined as $\mu_\alpha = \alpha\delta_{\delta_{\bar{\omega}}} + (1-\alpha)\delta_{\tilde{p}}$ and $\tilde{p} \in \Delta\Omega$ defined as in the statement of the proposition for a preferred state $\bar{\omega}$. Clearly $\alpha < \hat{p}(\bar{\omega})$ and $\mu_\alpha \in \Gamma(\hat{p})$ by construction. For each action f, $\mathbb{E}_{\tilde{p}}[u(f)] = \sum_{\omega \neq \bar{\omega}} \frac{\hat{p}(\omega)}{1-\alpha} u(f(\omega)) + \frac{\hat{p}(\bar{\omega})-\alpha}{1-\alpha} u(f(\bar{\omega})) = \frac{1}{1-\alpha}\mathbb{E}_{\hat{p}}[u(f)] - \frac{\alpha}{1-\alpha} u(f(\bar{\omega}))$. This means that

$$W(\delta_{\tilde{p}}, F) = \max_{f \in F} \frac{1}{1-\alpha} \left(\mathbb{E}_{\hat{p}}[u(f)] - \alpha u(f(\bar{\omega}))\right).$$

For α small enough, $W(\delta_{\tilde{p}}, F) = \frac{1}{1-\alpha}\left(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \alpha u(f^*_{F,\hat{p}}(\bar{\omega}))\right)$. By the condition $f^*_{F,\hat{p}}(\bar{\omega}) \leq \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$, $W(\delta_{\tilde{p}}, F) \geq \frac{1}{1-\alpha}\left(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \alpha\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]\right) = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$. Consider now

$$V(F|\mu_\alpha) = \alpha u(f(\bar{\omega})) + (1-\alpha)W(\delta_{\tilde{p}}, F) - \left(\alpha d(f^*_{F,\delta_{\bar{\omega}}}, f_{F,\delta_{\bar{\omega}}}, \delta_{\bar{\omega}}) + (1-\alpha)I(\delta_{\tilde{p}}, F)\right).$$

The fact that $W(\delta_{\tilde{p}}, F) \geq \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$, and the assumption that d is $\frac{\alpha}{1-\alpha}$-flat at $\hat{p}$ imply that $I(\delta_{\tilde{p}}, F) = d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$. Indeed, $I(\delta_{\tilde{p}}, F) = d(f^*_{F,\tilde{p}}, f_{F,\tilde{p}}, \tilde{p}) =$

$d(f^*_{F,\hat{p}}, f_{F,\tilde{p}}, \tilde{p}) = d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}) = d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$, because $\max_{\omega\in\Omega} |\hat{p}(\omega) - \tilde{p}(\omega)| \leq \frac{\alpha}{1-\alpha}$. Thus, $V(F|\mu_\alpha) > v(F, \hat{p})$ whenever $u(f(\bar{\omega})) - d(f^*_{F,\delta_{\bar{\omega}}}, f_{F,\delta_{\bar{\omega}}}, \delta_{\bar{\omega}}) > v(F, \hat{p})$. □

The test perfectly reveals a preferred state with a probability of $\alpha \leq \hat{p}(\bar{\omega})$, and is (almost) uninformative otherwise ($\tilde{p}$ is "close" to $\hat{p}$ for small α). Proposition B.1 can inform the design of medical tests. Suppose that, for a person with low blood pressure, consuming red meat has a sufficiently higher material value than not consuming it. A test that reveals low blood pressure with small probability and is otherwise uninformative is strictly valuable for a stereotypical masculine identity.

B.2. Identity uncertainty

In this section, I relax the assumption that the material value of an epistemic situation is the expected utility of the payoff-maximizing action. Instead, I consider a convex combination of the expected utility of the payoff-maximizing action and of the prescription. One interpretation is that the individual is uncertain about the strength of their "identity" motives (as in Bénabou and Tirole, 2011; Grossman and van der Weele, 2016). With some probability–which may depend on the belief–they will maximize the material utility in the second period, with complementary probability they will follow the prescription. The value of an epistemic situation (F, q) becomes

$$v_{\gamma_q}(F, q) = \gamma_q \mathbb{E}_q[u(f^*_{F,q})] + (1 - \gamma_q)\mathbb{E}_q[u(f_{F,q})] - \gamma_q d(f^*_{F,q}, f_{F,q}, q),$$

for some $\gamma_q \in [0, 1]$. The Identity model corresponds to $\gamma_q = 1$. The case $\gamma_q = 0$ represents the extreme case of an individual who always follows the prescriptions (e.g., an orthodox, who always follows the prescriptions of their religion). The ex-ante value of a menu F with information acquisition is

$$V_\gamma(F|\mu) = \int_{\Delta\Omega} v_{\gamma_q}(F, q) d\mu(q).$$

Under ignorance, it is $v_{\gamma_{\hat{p}}}(F, \hat{p}) = \gamma_{\hat{p}} \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + (1 - \gamma_{\hat{p}})\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] - \gamma_{\hat{p}} d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$. In general,

$$W(\mu, F) \geq \int_{\Delta\Omega} \gamma_q \mathbb{E}_q[u(f^*_{F,q})] + (1 - \gamma_q)\mathbb{E}_q[u(f_{F,q})] d\mu(q),$$

but $\int_{\Delta\Omega} \gamma_q d(f^*_{F,q}, f_{F,q}, q) d\mu(q) \leq I(\mu, F)$. Thus, the material value of F in the extended model is lower than in the baseline case, but the opposite holds for the psychological cost. Even though the model V_γ is more general than the baseline model, the two are *observationally equivalent* from the ex-ante point of view (i.e., without observing the second-period choice). Indeed, I can rewrite

$V_\gamma(F|\mu)$ as $V_\gamma(F|\mu) = W(\mu, F) - I_\gamma(\mu, F)$ where:

$$I_\gamma(\mu, F) = \int_{\Delta\Omega} \Big((1-\gamma_q)\Big(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})]\Big) + \gamma_q d(f^*_{F,q}, f_{F_q}, q)\Big)\, d\mu(q).$$

Thus, the extended model is equivalent to a baseline model in which the cost function is a convex combination of a "material component," measuring the foregone material value of following the prescription, and the pure psychological component. The intuition is that following the prescription in the second period eliminates the psychological cost, but it can be materially costly. To see this, consider the moral wiggle room example with the usual assumptions and with $\gamma_q = 0.5$ for $q \in \{\hat{p}, \delta_{\omega_1}, \delta_{\omega_2}\}$. Then, $v_{\gamma_{\hat{p}}}(a \cup b, \hat{p}) = \mathbb{E}_{\hat{p}}[u(a)] = 6$, whereas

$$\begin{aligned} V_\gamma(a \cup b|\bar{\mu}) =& 6 - \left[\frac{1}{2}\left(\frac{1}{2}(6-6) + \frac{1}{2}d(a, a, \delta_{\omega_1})\right) + \frac{1}{2}\left(\frac{1}{2}(6-5) + \frac{1}{2}d(a, b, \delta_{\omega_2})\right)\right] \\ &= 6 - \frac{1}{4} - \frac{1}{4}d(a, b, \delta_{\omega_2}). \end{aligned}$$

As in the baseline model, ignorance is strictly optimal. In the extended model, the desire to remain ignorant depends on the psychological cost and is reinforced by the possibly lower material payoff coming from following the prescription in state ω_2.

B.3. Commitment without uncertainty: costly exit in dictator games

In the Identity model, a preference for commitment is independent of the presence of uncertainty. Indeed, Proposition 3 holds even if $|\Omega| = 1$. Therefore, the model can rationalize the experimental evidence of Dana et al. (2006). In a laboratory experiment, the classic dictator game was modified to allow the dictators to either split \$10 or exit the game before playing and without informing the recipient. In the case of opting out, the dictator receives \$9 and the recipient \$0. One third of the experimental subjects opted out. Opting out is inconsistent with both purely altruistic and purely selfish preferences. In the former case, the available allocation $(9, 1)$ is better than opting out; in the latter, the allocation $(10, 0)$ is better than opting out. The dictators who exit seem to have a preference for avoiding a moral trade-off. Entering the game means facing $F = \{(x, 10-x), x \in \{0, \dots, 10\}\}$. I assume that the prescription is $(5, 5)$ in F. Then, the value of F is $v(F) = \max_{(x,10-x)\in F} u(x, 10-x) - d((x, 10-x)^*_F, (5, 5))$ (where I have suppressed the dependence on the degenerated prior). If the payoff-maximizing action is $(10, 0)$, then $v(F) = u(10, 0) - d((10, 0), (5, 5))$. Opting out implies commitment to $(9, 0)$, that has value $v((9, 0)) = u(9, 0)$. If the utility function is $u(x, y) = x$, then $v(F) = 10 - d((10, 0), (5, 5))$ and $u(9, 0) = 9$. Therefore, $v((9, 0)) \geq v(F)$ whenever $1 \leq d((10, 0), (5, 5))$.

Appendix C: Online Appendix

C.1. The envelope game

The Identity model rationalizes the experimental results of Serra-Garcia and Szech (2022). They designed a moral dilemma, called the Envelope Game, to study the "elasticity" of willful ignorance. Each subject received an envelope that with a probability of 0.5 contains a \$10 donation to a non-profit organization and with a probability of 0.5 is empty. The subject first chose whether to open the envelope or not. If they did not open the envelope, they decided between receiving a monetary amount or the uncertain envelope. If they opened the envelope, they learned its content and then chose between the content of the envelope or receiving a monetary amount. Table C.1 summarizes the payoffs of the game, where $m = o, c \in \mathbb{R}$ are monetary amounts that

	ω_1	ω_2
f_m	$(2.5 + m, 0)$	$(2.5 + m, 0)$
g_m	$(m, 0)$	$(m, 10)$

Table C.1. Actions and payoffs in Serra-Garcia and Szech (2022).

vary with the decision to open (o) or not (c) the envelope. In the experiment $o - c \in [-2, 2]$. Assume that $u(x, y) = x$ and that the prescription is to donate if the envelope contains the \$10 donation (in state ω_2), and to take the money if it is empty (in state ω_1). Information avoidance of the perfect information $\bar{\mu}$ corresponds to a preference for not opening the envelope, $v(f_c \cup g_c, \hat{p}) > V(f_o \cup g_o|\bar{\mu})$. Under the stated assumptions, $v(f_c \cup g_c, \hat{p}) = 2.5 + c - d(f_c, f_{f_c \cup g_c, \hat{p}}, \hat{p}) > 2.5 + o - 0.5\,[d(f_o, f_o, \delta_{\omega_1}) + d(f_o, g_o, \delta_{\omega_2})] = V(f_o \cup g_o|\bar{\mu})$ when

$$c - o > d(f_c, f_{f_c \cup g_c, \hat{p}}, \hat{p}) - 0.5d(f_o, g_o, \delta_{\omega_2}). \tag{C.1}$$

First, the larger the incentive to open the envelope o, the more difficult it is to satisfy the inequality (C.1) if the cost $d(f_o, g_o, \delta_{\omega_2})$ grows less than linearly in o. For instance, if $d = d_1$, $c - o > d_1(f_c, f_{f_c \cup g_c, \hat{p}}, \hat{p}) - 0.5$. Since the right-hand side of this inequality is independent of o, it becomes harder to satisfy as o increases, which is consistent with the results in Serra-Garcia and Szech (2022). A symmetric consideration applies to increasing the incentive to keep the envelope closed c. Lastly, if the left-hand side of (C.1) is negative, so that the incentive to open the envelope is larger than the incentive to leave the envelope closed, the inequality can still be satisfied if the cost $0.5d(f_o, g_o, \delta_{\omega_2})$ is large enough.

C.2. Local properties of the cost function d

From now on, I will occasionally use the following notation: $\sigma_F(q) = \max_{f \in F} \mathbb{E}_q[u(f)] = \mathbb{E}_q[u(f^*_{F,q})]$.

Information avoidance. In this section, I introduce local properties of the cost function d which help to refine some results in the main text. First, for any menu F and $q \in \operatorname{supp}\mu$, there exists $\lambda_q \in \mathbb{R}$ such that $d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}) \geq \lambda_q \left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})] - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right)$. If one can find a positive λ that satisfies these inequalities for all $q \in \operatorname{supp}\mu$, I say that d is λ-utility commensurable. Formally:

Definition C.1. The function d is λ-utility commensurable (at F and $\hat{p}$) if there is $\lambda \geq 0$ such that $d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}) \geq \lambda \left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})] - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right)$ for all $q \in \operatorname{supp}\mu$.

Intuitively, λ represents a common unit of measurement between abstract costs and material utility. The notion of λ-utility commensurability is "local," since λ may depend on F and $\hat{p}$. Moreover, multiple λ are possible, as the next example demonstrates:

Example C.1. Consider the moral wiggle room example with the usual assumptions about the prescriptions and the utility. There are two posteriors in the support of μ, δ_{ω_1} and δ_{ω_2}. For the posterior δ_{ω_1}, $d(a, a, \delta_{\omega_1}) - d(a, a, \hat{p}) = 0 \geq \lambda_{\delta_{\omega_1}}(6 - 6 - 6 + 6) = 0$, which is true for any $\lambda_{\delta_{\omega_1}}$. For the posterior δ_{ω_2}, $d(a, b, \delta_{\omega_2}) - d(a, a, \hat{p}) \geq \lambda_{\delta_{\omega_2}}(6 - 5 - 6 + 6)$, which holds for any $0 \leq \lambda_{\delta_{\omega_2}} \leq d(a, b, \delta_{\omega_2})$. It follows that d is λ-utility commensurable for any $\lambda \in [0, d(a, b, \delta_{\omega_2})]$. ■

The next proposition provides a sufficient condition for having a λ-utility commensurable cost. Given a menu F and $\hat{p} \in \Delta\Omega$, I define for any $q \in \operatorname{supp}\mu$, $d_q = d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$ and $x_q = \mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})] - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]$. By definition, the cost d is λ-utility commensurable (at F and $\hat{p}$) if $d_q \geq \lambda x_q$ for some $\lambda \geq 0$ and all $q \in \operatorname{supp}\mu$. Let $D_+ = \{d_q : d_q > 0\}$, $D_- = \{d_q : d_q < 0\}$ and $X_+ = \{x_q : x_q > 0\}$ and $X_- = \{d_x : d_x < 0\}$. Lastly, define $F_+ = \left\{\frac{d_q}{x_q} : d_q \in D_+,\ x_q \in X_+\right\}$ and $F_- = \left\{\frac{d_q}{x_q} : d_q \in D_-,\ x_q \in X_-\right\}$.

Proposition C.1 (Sufficient conditions for commensurability). *The cost d is λ-utility commensurable for all (if any) λ satisfying* $\max_{F_+} \frac{d_q}{x_q} \geq \lambda \geq \min_{F_-} \frac{d_q}{x_q}$.

The proof is immediate. Information avoidance for a menu F is equivalent to the cost of information $(I(\mu, F) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}))$ being strictly greater than the "material value of information" $(W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]) \geq 0)$. Endowed with the definition of λ-utility commensurability, the following result gives sufficient conditions for information avoidance.

Proposition C.2 (Information avoidance with commensurable cost). *For all menus F, if d is λ-utility commensurable then:*

1. *If $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] > \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$ and $\lambda \geq 1$, then there is information avoidance for F.*
2. *If $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] = \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$, $W(\mu, F) > \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$ and $\lambda > 1$, then there is information avoidance for F.*

Proof of Proposition C.2. For case 1, I have the following sequence of inequalities:

$$
\begin{aligned}
&\int_{\Delta\Omega} \left(d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})\right) d\mu(q) \\
&\geq \lambda \int_{\Delta\Omega} \left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})]\right) d\mu(q) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] \\
&> \lambda \left(W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]\right) \\
&\geq W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})],
\end{aligned}
$$

where the first inequality follows from the fact that d is λ-utility commensurable. The second inequality follows from $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] > \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$ and the last from $\lambda \geq 1$ and $W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] \geq 0$. Rearranging gives the result.

For case 2, it holds that

$$
\begin{aligned}
&\int_{\Delta\Omega} \left(d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})\right) d\mu(q) \\
&\geq \lambda \int_{\Delta\Omega} \left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})]\right) d\mu(q) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] \\
&= \lambda \left(W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]\right) \\
&> W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})],
\end{aligned}
$$

where the first inequality follows from the fact that d is λ-utility commensurable. The equality follows from $\int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q) = \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]$ and the strict inequality from $\lambda > 1$ and $W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] > 0$. Rearranging gives the result. □

When the cost is λ-utility commensurable, the difference $I(\mu, F) - d(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p})$ is proportional to $W(\mu, F) - \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]$. Acquiring information increases the material value of F, but also its cost (proportionally to $W(\mu, F) - \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$). If the "marginal cost" of information λ is sufficiently high, ignorance is optimal. Note that, if there is an absolute prescription in F, the condition in point $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] = \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$ is automatically satisfied. In this case, information avoidance for F requires only a sufficiently high marginal cost. The notion of λ-utility commensurability is related to the subdifferential[19] of φ when $d = d_e$:

Corollary C.1 (Cognitive dissonance and information avoidance). *Suppose that* $d = d_e$. *If* $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] > \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$ *and* $\lambda \geq 1$ *for some* $\lambda \in \partial\varphi\left(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right)$, *then there is information avoidance for* F. *If* $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] = \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$, $W(\mu, F) > \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$ *and* $\lambda > 1$ *for some* $\lambda \in \partial\varphi\left(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right)$, *then there is information avoidance for* F.

Proof of Corollary C.1. By convexity of φ, for any $q \in \operatorname{supp}\mu$ it holds that

$$d_e(f^*_{F,q}, f_{F,q}, q) - d_e(f^*_{F,\hat{p}}, f_{F,\hat{p}}, \hat{p}) \geq \lambda\left[\sigma_F(q) - \mathbb{E}_q[u(f_{F,q})] - \sigma_F(\hat{p}) + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right],$$

for some $\lambda \in \partial\varphi(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})])$. Integrating with respect to μ gives

$$\begin{aligned}
&\int_{\Delta\Omega} \varphi(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})])d\mu(q) - \varphi(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]) \\
&\geq \lambda\left[W(\mu, F) - \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right] \\
&> \lambda\left[W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]\right] \\
&\geq W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})],
\end{aligned}$$

where the first inequality follows from the monotonicity of the integral, the strict inequality from the condition $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] > \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$, and the last inequality from $\lambda \geq 1$ and $W(\mu, F) \geq \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$. Rearranging gives the result. For the second case:

$$\begin{aligned}
&\int_{\Delta\Omega} \varphi(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})])d\mu(q) - \varphi(\mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] - \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]) \\
&\geq \lambda\left[W(\mu, F) - \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})] + \mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})]\right] \\
&= \lambda\left[W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]\right] \\
&> W(\mu, F) - \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})],
\end{aligned}$$

where the first inequality follows from the monotonicity of the integral, the equality from the condition $\mathbb{E}_{\hat{p}}[u(f_{F,\hat{p}})] = \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$, and the last inequality from $\lambda > 1$ and $W(\mu, F) > \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$. □

Avoidance of situations. Given two menus F and G, a preference for commitment to F is equivalent to the cost of flexibility $(I(\mu, F \cup G) - I(\mu, F))$

19. The subdifferential of $\varphi : \mathbb{R} \to \mathbb{R} \cup \infty$ at $x \in \mathbb{R}$ is the set $\partial\varphi(x) = \{\lambda \in \mathbb{R} : \varphi(y) - \varphi(x) \geq \lambda(y - x),\ \forall y \in \mathbb{R}\}$. It is convex, closed and non-empty on the relative interior of the domain of f.

being strictly greater than the "material value of flexibility" ($W(\mu, F \cup G) - W(\mu, F)$). In this section, I introduce local properties of the cost function d that are analogous to λ-utility commensurability. These properties provide parametric sufficient conditions for a preference for commitment.

Definition C.2. The function d is λ_q-utility commensurable (at F, G) if there is a $\lambda_q \geq 0$ such that

$$d(f^*_{F,q}, f_{F,q}, q) - d(f^*_{G,q}, f_{G,q}, q) \geq \lambda_q \left(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})] - \mathbb{E}_q[u(f^*_{G,q})] + \mathbb{E}_q[u(f_{G,q})]\right).$$

The notion of λ_q-utility commensurability is local, since it depends on the belief q and the menus F and G. Commensurability is linked to regularity of the cost function.

Fact C.1. *If d is λ_q-utility commensurable (at $F \cup G, F$) for all $q \in \operatorname{supp}\mu$ and $q = \hat{p}$, then it is regular.*

Proof of Fact C.1. Suppose that $F \cup G$ is prescriptively equivalent to F and d is λ_q-utility commensurable for all $q \in \operatorname{supp}\mu$ and $q = \hat{p}$, then:

$$\begin{aligned} d(f^*_{F\cup G,q}, f, q) - d(f^*_{F,q}, f, q) &\geq \lambda_q \left(\mathbb{E}_q[u(f^*_{F\cup G,q})] - \mathbb{E}_q[u(f)] - \mathbb{E}_q[u(f^*_{F,q})] + \mathbb{E}_q[u(f)]\right) \\ &= \lambda_q \left(\mathbb{E}_q[u(f^*_{F\cup G,q})] - \mathbb{E}_q[u(f^*_{F,q})]\right) \geq 0, \end{aligned}$$

implying regularity. □

The next result gives a sufficient condition for a preference for commitment when the cost d is λ_q-utility commensurable for all $q \in \operatorname{supp}\mu$.

Proposition C.3 (Avoidance of situations with commensurable cost). *Suppose that $F \cup G$ is prescriptively equivalent to F and d is λ_q-utility commensurable (at $F \cup G, F$) for all $q \in \operatorname{supp}\mu$. If $\lambda_* \geq 1$, where $\lambda_* = \min_{q \in \operatorname{supp}\mu} \lambda_q$, then $V(F|\mu) \geq V(F \cup G|\mu)$.*

Proof of Proposition C.3. I have the following sequence of inequalities:

$$\begin{aligned} &\int_{\Delta\Omega} \left(d(f^*_{F\cup G,q}, f_{F\cup G,q}, q) - d(f^*_{F,q}, f_{F,q}, q)\right) d\mu(q) \\ &\geq \int_{\Delta\Omega} \lambda_q \left(\sigma_{F\cup G}(q) - \mathbb{E}_q[u(f_{F\cup G,q})] - \sigma_F(q) + \mathbb{E}_q[u(f_{F,q})]\right) d\mu(q) \\ &= \int_{\Delta\Omega} \lambda_q \left(\sigma_{F\cup G}(q) - \sigma_F(q)\right) d\mu(q) \\ &\geq \lambda_* \left(W(\mu, F \cup G) - W(\mu, F)\right) \\ &\geq W(\mu, F \cup G) - W(\mu, F), \end{aligned}$$

where the first inequality follows from the fact that d is λ_q-utility commensurable (at $F \cup G, F$). The equality follows from the fact that $F \cup G$ is prescriptively equivalent to F (hence $\int_{\Delta\Omega} \mathbb{E}_q[u(f_{F\cup G,q})]d\mu(q) = \int_{\Delta\Omega} \mathbb{E}_q[u(f_{F,q})]d\mu(q)$). The second inequality follows from $\sigma_{F\cup G}(q) \geq \sigma_F(q)$ for all q, and the last inequality from $\lambda_* \geq 1$ and $W(\mu, F \cup G) \geq W(\mu, F)$. Rearranging gives the result. □

Lastly, for the case $d = d_e$, a preference for commitment depends on the slope of φ.

Corollary C.2 (Cognitive dissonance and avoidance of situations). *Suppose that $d = d_e$ and define $\lambda_* = \min_{q\in\operatorname{supp}\mu} \lambda_q$ for $\lambda_q \in \partial\varphi(\mathbb{E}_q[u(f^*_{F,q})] - \mathbb{E}_q[u(f_{F,q})])$. If $F \cup G$ is prescriptively equivalent to F and $\lambda_* \geq 1$, then $V(F|\mu) \geq V(F \cup G|\mu)$.*

Proof of Corollary C.2. Since φ is convex, given $q \in \operatorname{supp}\mu$, there is $\lambda_q \in \partial\varphi(\sigma_F(q) - \mathbb{E}_q[u(f_{F,q})])$ such that

$$\begin{aligned}
&\varphi(\sigma_{F\cup G}(q) - \mathbb{E}_q[u(f_{F\cup G,q})]) - \varphi(\sigma_F(q) - \mathbb{E}_q[u(f_{F,q})]) \\
&\geq \lambda_q \left[\sigma_{F\cup G}(q) - \mathbb{E}_q[u(f_{F\cup G,q})] - \sigma_F(q) + \mathbb{E}_q[u(f_{F,q})]\right] \\
&= \lambda_q \left[\sigma_{F\cup G}(q) - \sigma_F(q)\right],
\end{aligned}$$

where the equality follows from the fact that $F \cup G$ is prescriptively equivalent to F. Integrating with respect to μ gives

$$\begin{aligned}
&\int_{\Delta\Omega} \Big(\varphi(\sigma_{F\cup G}(q) - \mathbb{E}_q[u(f_{F\cup G,q})]) - \varphi(\sigma_F(q) - \mathbb{E}_q[u(f_{F,q})])\Big)\, d\mu(q) \\
&\geq \int_{\Delta\Omega} \lambda_q \left(\sigma_{F\cup G}(q) - \sigma_F(q)\right) d\mu(q) \\
&\geq \lambda_* \int_{\Delta\Omega} \left(\sigma_{F\cup G}(q) - \sigma_F(q)\right) d\mu(q) \\
&= \lambda_* \left(W(\mu, F \cup G) - W(\mu, F)\right) \\
&\geq W(\mu, F \cup G) - W(\mu, F),
\end{aligned}$$

where the second inequality follows from the fact that $\sigma_{F\cup G}(q) \geq \sigma_F(q)$ for all beliefs q. The last inequality from $\lambda_* \geq 1$ and $W(\mu, F \cup G) \geq W(\mu, F)$. Rearranging gives the result. □

C.3. Meta-prescriptions: "acting white"

In this section, I complement the extension of my model outlined in Section 6. I discuss meta-prescriptions about future opportunities. For example, the norms prohibiting married women from entering the labor market (e.g., Bertrand et al., 2015), or black students from accumulating human capital so as to avoid "acting white" (e.g., Austen-Smith and Fryer Jr, 2005). Suppose there are two menus of activities F and B. The Identity model assigns value to the pairs $F_\mu, F_{\delta_{\hat{p}}}, B_\mu, B_{\delta_{\hat{p}}}$. The extended model assigns value to generalized menus, e.g., $\mathbb{G} = \{F_\mu, B_\mu\}$. A meta-prescription of not acting white in $\mathbb{G}$ is B_μ. Assume that accumulating human capital means facing the generalized menu $\mathbb{F} = \{F_\mu, B_\mu, F_{\delta_{\hat{p}}}, B_{\delta_{\hat{p}}}\}$. As for the interpretation, $\mathbb{F}$ means having the future possibility to select any activity in F and B (plus acquiring or rejecting information). If B_μ is the meta-prescription in $\mathbb{F}$,

$$\hat{V}(\mathbb{F}) = \max\{V(F|\mu), V(B|\mu), v(F,\hat{p}), v(B,\hat{p})\} - D(H^*_{\mathbb{F}}, B_\mu).$$

For simplicity, I assume that there are no identity trade-offs in the second stage. This implies $V(F|\mu) = W(\mu, F)$, $V(B|\mu) = W(\mu, B)$, $v(F,\hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*_{F,\hat{p}})]$ and $v(B,\hat{p}) = \mathbb{E}_{\hat{p}}[u(f^*_{B,\hat{p}})]$. If the "white" activities are materially superior to the "black" activities (e.g., if $B \subseteq F$), the value of accumulating human capital is $\hat{V}(\mathbb{F}) = W(\mu, F) - D(F_\mu, B_\mu)$, whereas the value of not accumulating human capital is $\hat{V}(B_\mu \cup B_{\delta_{\hat{p}}}) = W(\mu, B)$. The interpretation is that only the activities in B will be available in the future. If the psychological cost $D(F_\mu, B_\mu)$ of violating the meta-prescription is sufficiently high, committing to B_μ becomes optimal. The identity sanctions the mere act of acquiring certain skills that could potentially lead to an activity in F. The choice of not accumulating human capital does not require a public signalling role of behavior, but simply a psychological cost; and this is regardless of the activity eventually chosen by the student.

References

Abrams, Dominic and Michael A. Hogg (2006). *Social identifications: A social psychology of intergroup relations and group processes.* Routledge.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and identity." *The Quarterly Journal of Economics*, 115(3), 715–753.

Akerlof, George A. and Rachel E. Kranton (2005). "Identity and the Economics of Organizations." *Journal of Economic Perspectives*, 19(1), 9–32.

Akerlof, George A. and Rachel E. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being.* Princeton University Press, URL http://www.jstor.org/stable/j.ctt7rqsp.

Alaoui, Larbi (2016). "Information avoidance and the preservation of self-image." *Mimeo.*

Andersson, Per A., Arvid Erlandsson, and Daniel Västfjäll (2022). "Norm avoiders: The effect of optional descriptive norms on charitable donations." *Journal of Behavioral Decision Making*, 35(1), e2244.

Andreoni, James, Justin M. Rao, and Hannah Trachtman (2017). "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." *Journal of Political Economy*, 125(3), 625–653.

Atkin, David, Eve Colson-Sihra, and Moses Shayo (2021). "How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption." *Journal of Political Economy*, 129(4), 1193–1251.

Austen-Smith, David and Roland G. Fryer Jr (2005). "An economic analysis of "acting white"." *The Quarterly Journal of Economics*, 120(2), 551–583.

Ballester, Miguel A. and Irem Bozbay (2021). "Revealed identities." *Mimeo.*

Bartling, Björn, Florian Engl, and Roberto A. Weber (2014). "Does willful ignorance deflect punishment?–An experimental study." *European Economic Review*, 70, 512–524.

Bašić, Zvonimir and Eugenio Verrina (2021). "Personal norms—and not only social norms—shape economic behavior." *MPI Collective Goods Discussion Paper*, (2020/25).

Bénabou, Roland and Jean Tirole (2011). "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126(2), 805–855.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland (2010). "Social identity and preferences." *American Economic Review*, 100(4), 1913–28.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan (2015). "Gender identity and relative income within households." *The Quarterly Journal of Economics*, 130(2), 571–614.

Blackwell, David (1953). "Equivalent Comparisons of Experiments." *Ann. Math. Statist.*, 24(2), 265–272.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini (2021). "Identity, beliefs, and political conflict." *The Quarterly Journal of Economics*, 136(4), 2371–2411.

Brocas, Isabelle and Juan D. Carrillo (2004). "Entrepreneurial boldness and excessive investment." *Journal of Economics & Management Strategy*, 13(2), 321–350.

Courtenay, Will H. (2000). "Constructions of masculinity and their influence on men's well-being: a theory of gender and health." *Social Science & Medicine*, 50(10), 1385–1401.

D'Acunto, Francesco (2019). "Identity and choice under risk." *Available at SSRN 3466626.*

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes (2006). "What you don't know won't hurt me: Costly (but quiet) exit in dictator games." *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33(1), 67–80.

De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek (2017). "Rationally inattentive preferences and hidden information costs." *Theoretical Economics*, 12(2), 621–654.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012). "Testing for altruism and social pressure in charitable giving." *The Quarterly Journal of Economics*, 127(1), 1–56.

Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014). “A theory of subjective learning.” *Journal of Economic Theory*, 153(0), 287 – 312.

Dillenberger, David and Philipp Sadowski (2012). “Ashamed to be selfish.” *Theoretical Economics*, 7(1), 99–124.

Dohmen, Thomas and Armin Falk (2011). “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender.” *American Economic Review*, 101(2), 556–90.

Ellingsen, Tore and Erik Mohlin (2019). “Decency.” Tech. rep.

Epstein, Larry G (2006). “An axiomatic model of non-Bayesian updating.” *The Review of Economic Studies*, 73(2), 413–436.

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni (2008). “Non-Bayesian updating: a theoretical framework.” *Theoretical Economics*, 3(2), 193–229.

Feiler, Lauren (2014). “Testing models of information avoidance with binary choice dictator games.” *Journal of Economic Psychology*, 45, 253–267.

Festinger, Leon (1962). “Cognitive dissonance.” *Scientific American*, 207(4), 93–106.

Fong, Christina M. and Felix Oberholzer-Gee (2011). “Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor.” *Journal of Public Economics*, 95(5), 436–444. Charitable Giving and Fundraising Special Issue.

Gilboa, Ithzak, Stefania Minardi, and Fang Wang (2022). “Meaning and discontinuity in consumer choice.” *Mimeo.*

Golman, Russell, David Hagmann, and George Loewenstein (2017). “Information avoidance.” *Journal of Economic Literature*, 55(1), 96–135.

Grossman, Zachary (2014). “Strategic Ignorance and the Robustness of Social Preferences.” *Management Science*, 60(11), 2659–2665.

Grossman, Zachary and Joël J. van der Weele (2016). “Self-image and willful ignorance in social decisions.” *Journal of the European Economic Association*, 15(1), 173–217.

Hertwig, Ralph and Christoph Engel (2016). “Homo Ignorans: Deliberately Choosing Not to Know.” *Perspectives on Psychological Science*, 11(3), 359–372. PMID: 27217249.

Hoffman, Moshe, Erez Yoeli, and Martin A. Nowak (2015). “Cooperate without looking: Why we care what people think and not just what they do.” *Proceedings of the National Academy of Sciences*, 112(6), 1727–1732.

Kamenica, Emir and Matthew Gentzkow (2011). “Bayesian persuasion.” *American Economic Review*, 101(6), 2590–2615.

Kandul, Serhiy and Ilana Ritov (2017). “Close your eyes and be nice: Deliberate ignorance behind pro-social choices.” *Economics Letters*, 153, 54–56.

Konow, James (2000). “Fair shares: Accountability and cognitive dissonance in allocation decisions.” *American Economic Review*, 90(4), 1072–1091.

Köszegi, Botond (2006). “Ego utility, overconfidence, and task choice.” *Journal of the European Economic Association*, 4(4), 673–707.

Kranton, Rachel E. (2016). "Identity Economics 2016: Where Do Social Distinctions and Norms Come From?" *American Economic Review*, 106(5), 405–09.

Krupka, Erin L. and Roberto A. Weber (2013). "Identifying social norms using coordination games: why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3), 495–524.

Larson, Tara and C. Monica Capra (2009). "Exploiting moral wiggle room: Illusory preference for fairness? A comment." *Judgment and Decision Making*, 4(6), 467.

Lipnowski, Elliot and Laurent Mathevet (2018). "Disclosure to a psychological audience." *American Economic Journal: Microeconomics*, 10(4), 67–93.

Loewenstein, George and Andras Molnar (2018). "The renaissance of belief-based utility in economics." *Nature Human Behaviour*, 2(3), 166–167.

Mahalik, James R., Shaun M. Burns, and Matthew Syzdek (2007). "Masculinity and perceived normative health behaviors as predictors of men's health behaviors." *Social Science & Medicine*, 64(11), 2201–2209.

Matthey, Astrid and Tobias Regner (2011). "Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior." *Games*, 2(1), 114–135.

Momsen, Katharina and Markus Ohndorf (2020). "Expressive Voting vs. Self-Serving Ignorance." Tech. rep., Working Papers in Economics and Statistics.

Niederle, Muriel and Lise Vesterlund (2007). "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Niehaus, Paul (2014). "A theory of good intentions." *Mimeo.*

Nyborg, Karine (2011). "I don't want to hear about it: Rational ignorance among duty-oriented consumers." *Journal of Economic Behavior & Organization*, 79(3), 263–274.

Oh, Suanna (2021). "Does identity affect labor supply?" *Available at SSRN 3998025.*

Pennesi, Daniele (2015). "Costly information acquisition and the temporal resolution of uncertainty." *Journal of Mathematical Economics*, 60, 115–122.

Piermont, Evan (2019). "Images and norms." *Mimeo.*

Pomatto, Luciano, Philipp Strack, and Omer Tamuz (2023). "The Cost of Information: The Case of Constant Marginal Costs." *American Economic Review*, p. Forthcoming.

Rabin, Matthew (1995). "Moral preferences, moral constraints, and self-serving biases." *Mimeo.*

Rudman, Laurie A. and Kris Mescher (2013). "Penalizing Men Who Request a Family Leave: Is Flexibility Stigma a Femininity Stigma?" *Journal of Social Issues*, 69(2), 322–340.

Schwartz, Daniel, Elizabeth A. Keenan, Alex Imas, and Ayelet Gneezy (2021). "Opting-in to prosocial incentives." *Organizational Behavior and Human Decision Processes*, 163, 132–141. Nudges and Choice Architecture in

Organizations.

Serra-Garcia, Marta and Nora Szech (2022). "The (In)Elasticity of Moral Ignorance." *Management Science*, 68(7), 4815–4834.

Sims, C. A. (2003). "Implications of rational inattention." *Journal of Monetary Economics*, 50(3), 665–690.

Spiekermann, Kai and Arne Weiss (2016). "Objective and subjective compliance: A norm-based explanation of 'moral wiggle room'." *Games and Economic Behavior*, 96, 170–183.

Stets, Jan E. and Peter J. Burke (2000). "Identity theory and social identity theory." *Social Psychology Quarterly*, pp. 224–237.

Sunstein, Cass R (2020). *Too Much Information: Understanding What You Don't Want to Know*. MIT Press.

Trimmer, Pete, Richard McElreath, Sarah Auster, G. Brown, Jason Dana, Gerd Gigerenzer, Russell Golman, Christian Hilbe, Anne Kandler, Yaakov Kareev, et al. (2020). "The zoo of models of deliberate ignorance." In *Deliberate Ignorance: Choosing Not to Know*, vol. 29.

van Veldhuizen, Roel (2022). "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness?" *Journal of the European Economic Association*, 20(4), 1595–1618.

Vandello, Joseph A., Vanessa E. Hettinger, Jennifer K. Bosson, and Jasmine Siddiqi (2013). "When Equal Isn't Really Equal: The Masculine Dilemma of Seeking Work Flexibility." *Journal of Social Issues*, 69(2), 303–321.

Williams, Joan C., Mary Blair-Loy, and Jennifer L. Berdahl (2013). "Cultural Schemas, Social Class, and the Flexibility Stigma." *Journal of Social Issues*, 69(2), 209–234.