SOCIAL INFLUENCES, IDENTITIES, INTENTIONS:

ESSAYS ON FISHBEIN AND AJZEN'S REASONED ACTION APPROACH TO PREDICTING AND EXPLAINING BEHAVIOR

Marie Clare Barnett

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Dr. Cristina Bicchieri

S.J.P. Harvie Professor of Social Thought and Comparative Ethics

Graduate Group Chairperson

Dr. Errol Lord, Associate Professor of Philosophy

Dissertation Committee

Dr. Lisa Miracchi Titus, Assistant Professor of Philosophy, University of Denver

Dr. Jennifer Morton, Presidential Penn Compact Associate Professor of Philosophy

For Sophie

and in loving memory of Oliver

ACKNOWLEDGEMENTS

I am deeply indebted to my dissertation committee. Were it not for the advice and support of my advisor, Cristina Bicchieri, I would not have embarked upon this project, let alone completed it. I am immensely grateful to Lisa Miracchi Titus for her invaluable feedback and abundant generosity, and to Jennifer Morton for her insight and her kindness. Over the years, I have also been grateful to receive valuable feedback and advice from Jonathan Baron, Zoltan Domotor, Errol Lord, and Michael Weisberg.

I would like to extend my sincere thanks to the wonderful community that is the Department of Philosophy at Penn. I am grateful to everyone from whom I had the privilege to learn, and in particular to Karen Detlefsen and Gary Hatfield, who helped me grow both as a student and as a teacher. To my fellow graduate students, thank you for your insights, your camaraderie, and – most of all – your friendship. Special thanks to my cohort, Devin Curry, Daniel Fryer, Kurt Gerry, Nabeel Hamid, Karen Kovaka, and Jordan Taylor, and to Justin Bernstein, Louise Daoust, Lindsey Fiorelli, Allauren Forbes, Dylan Manson, Pierce Randall, Carlos Santana, and Rob Willison.

As my career as a student draws to a close, I would be remiss if I did not thank my undergraduate advisor, Duncan MacIntosh, and my master's supervisor, William Harper, who set me on my academic path. I am also very thankful to Brit Shields and Richard Shell, for giving me opportunities to broaden my horizons and for helping to make the completion of this dissertation possible.

I am deeply grateful to my brother, Arden, and to my parents, Anne and Grahame, for their unwavering love and support. Words cannot express my gratitude to Raffi, for everything.

ABSTRACT

SOCIAL INFLUENCES, IDENTITIES, INTENTIONS:

ESSAYS ON FISHBEIN AND AJZEN'S REASONED ACTION APPROACH TO PREDICTING AND EXPLAINING BEHAVIOR

Marie Clare Barnett

Cristina Bicchieri

This dissertation consists of three essays on the Fishbein-Ajzen model of human behavior, the most famous iteration of which is the theory of planned behavior. In chapter one, I argue that Fishbein and Ajzen's definitions of the attitudinal and normative components of their theory result in unnecessary and undesirable overlap between the two. In chapter two, I discuss Fishbein and Ajzen's assessment of self-identity as a potential addition to their theory and argue – contra Fishbein and Ajzen – that many of the measures of self-identity that are typically used are not likely to measure the theory's existing determinants of behavior, as opposed to a truly separate construct. In chapter three, I argue that Fishbein and Ajzen's definition of intention as behavioral expectation is problematic because 1) it does not match our commonsense notion of intention, and 2) it is the commonsense notion that is best suited to fill the role of "intention" within the context of the theory.

TABLE OF CONTENTS

Introduction

Since its debut in the late 1970s, the theory of planned behavior (TPB) has become one of the most prominent theoretical frameworks used by researchers who wish to understand and predict human behavior, providing the foundation for thousands of empirical studies (Bosnjak et al. 2020, 354). The dominant approach guiding research in the domain of health-related behavior (Sniehotta et. al. 2014, 1), it is well known for frequently outperforming its main rival, the Health Belief Model (see for example Şimşekoğlu and Lajunen's 2008 study of seat belt use, Gerend and Shepherd's 2012 study of HPV vaccination, and Montanaro and Bryan's 2014 study of condom use).

However, despite these successes, the theory continues to receive intense scrutiny. The theory's completeness is often questioned; extensions are frequently proposed (Fishbein and Ajzen 2010, 406). In recent years, some have even called for the theory to be retired (Sniehotta et. al. 2014). Nevertheless, the TPB remains one of the most used – and most discussed – theories in the social and behavioral sciences. Since 2018, it has been cited in a new publication approximately once an hour, according to statistics collected by Google Scholar. Though these statistics are not perfectly reliable, they give an impression of the theory's colossal reach. The TPB's successor is the latest iteration of the Fishbein-Ajzen model: the reasoned action approach (RAA). The purpose of this dissertation is to evaluate the RAA from a philosophical perspective, with an eye to improving the predictive and explanatory value of the model.

A Brief History of the RAA

The TPB evolved out of Martin Fishbein and Icek Ajzen's earlier theory, the theory of reasoned action (TRA). The TRA originated as an expectancy-value model proposed by Fishbein in the late 1960's. According to Fishbein's original theory – which was an extension of Don E. Dulany's (1968) theory of propositional control – "intentions are the immediate antecedents of behavior," and "intentions [...] are a function of attitude toward the behavior and the sum of normative beliefs weighted by motivation to comply" (Fishbein & Ajzen [hereafter F&A] 2010, 17-18).

Unlike the "minitheories" that were popular at the time of its publication, the TRA is a general model, applicable to all kinds of behavior. In the TRA, as in later versions of the theory, attitudes are assumed to be determined by behavioral beliefs (i.e. outcome expectancies) (F&A 2010, 18). A normative belief is understood as a "person's belief that [a] reference group or individual [...] thinks he should or should not perform [a] behavior" (F&A 1975, 302). Fishbein and Ajzen jointly authored two books on the theory; in the second, which was published in 1980, they began referring to the theory as the theory of reasoned action (F&A 2010, 218).

Fishbein and Ajzen went on to work separately. Ajzen continued to refine the theory, adding the construct of perceived behavioral control in order to accommodate behaviors that are not under complete volitional control. The theory thus extended is the TPB. The relationships between the main constructs are expressed in the following equation, where B_t represents "the strength of a person's attempt to perform a behavior" (Ajzen 1985, 30), I_t represents a person's "intention to try" to perform the behavior, A_t is

their "attitude toward trying", SN_t is the "subjective norm with regard to trying", and W_1 and W_2 are empirically determined weights (Ajzen 1985, 31):

$$B_t \sim I_t \propto [w_1A_t + w_2SN_t]$$

Ajzen explains that "the wavy line between B_t and I_t indicates that expressed intentions to try performing a behavior can change before the behavioral attempt is observed" (ibid.). Perceived behavioral control does not appear in the above equation because at the time, Ajzen conceived of perceived behavioral control as influencing a person's attitude toward trying to perform a behavior. Attitude toward trying was assumed to be directly proportional to attitude toward a successful attempt multiplied by the subjective probability of success, plus attitude toward a failed attempt multiplied by the subjective probability of failure ($A_t \propto [P_sA_s + P_fA_f]$)[1] (Ajzen 1985, 31).[2]

After participating in a National Institute of Mental Health workshop alongside Albert Bandura, Marshall Becker, Frederick Kanfer, and Harry Triandis in 1991 (F&A 2010, 18), Fishbein presented his own updated model in the late 1990s. This model, called the Integrative Model (IM), incorporated Bandura's concept of self-efficacy (Fishbein 2000, 274) and included an expanded normative component which accounted for the influence of descriptive as well as injunctive norms (F&A 2010, 19).

---

[1] A_t = attitude toward trying
P_s = subjective probability of success
A_s = attitude toward success
P_f = subjective probability of failure
A_f = attitude toward failure
$P_s + P_f = 1.0$

[2] Although Ajzen also expresses the view that "a behavioral intention can best be interpreted as an intention to try performing a certain behavior" (Ajzen 1988, 110) in his 1988 book, *Attitudes, Personality, and Behavior*, he does not offer the same analysis of perceived behavioral control as influencing intention to try in that work, nor does he present an equation describing the relationship between behavioral intention, perceived norm, and attitude.

In the 2000s, Fishbein and Ajzen joined forces once more, combining their theories to create the most recent version of their theoretical approach: the RAA. Like Ajzen's TPB, the RAA includes the concept of perceived behavioral control, and like Fishbein's IM, it includes an expanded normative component – termed "perceived norm" – which incorporates both injunctive and descriptive norms. According to the RAA, "in combination, attitude toward the behavior, perceived norm, and perception of behavioral control lead to the formation of a behavioral intention […] [T]he more favorable the attitude and perceived norm, and the greater the perceived behavioral control, the stronger should be the person's intention to perform the behavior in question" (F&A 2010, 21).

In the RAA, attitude, perceived norm, and perceived behavioral control are the three immediate antecedents of intention. According to the theory, all other potential influences on intention – e.g. personal standards and anticipated emotions – are indirect influences which are mediated by these factors. For example, your anticipated emotional reaction to the performance of a behavior (e.g. the anxiety you might feel during a dental exam) contributes to your attitude toward the behavior (i.e. your overall evaluation of the behavior), which – in combination with perceived norm and perceived behavioral control – directly leads to your intention to perform the behavior (or lack thereof). Similarly, your beliefs about the specific normative expectations and behaviors of particular others influence your perceived norm with regards to a behavior, which in turn directly influences your intention.

Key Definitions and Concepts

In *Predicting and Changing Behavior: The Reasoned Action Approach,* F&A offer the following definitions:

> *Attitude*: "the evaluation of an object, concept, or behavior, along a dimension of favor or disfavor, good or bad, like or dislike" (F&A 2010, 78). In the context of the RAA, the objects of attitudes are behaviors.
>
> *Perceived norm (alternatively: perceived social pressure)*: "the total social pressure experienced with respect to a given behavior", where this is understood as including pressure deriving from:
>
> > 1) *injunctive norms:* "perceptions regarding what should or ought to be done", and
> >
> > 2) *descriptive norms*: "perceptions that others are or are not performing the behavior in question" (F&A 2010, 131).
>
> *Perceived behavioral control*: "the extent to which people believe that they are capable of, or have control over, performing a given behavior" (F&A 2010, 155).
>
> *Intention*: "the subjective probability of performing a behavior" (F&A 2010, 40), or "readiness to engage in a behavior, a construct that incorporates such concepts as willingness, behavioral expectation, and trying" (F&A 2010, 43).

Another key concept is the *principle of compatibility*, according to which "an intention is compatible with a behavior if both are measured at the same level of generality or specificity—that is, if the measure of intention involves exactly the same action, target, context, and time elements as the measure of behavior" (F&A 2010, 44). The purpose of this principle is to ensure a good match between predictor and behavior – something which is necessary because, for example, a person's positive attitude toward "going to the movies" will not necessarily incline them toward engaging in the behavior "watching a horror movie at the theater at least twice a week" (perhaps they only like comedies). The principle has implications not only for how behaviors and intentions are measured, but

also for how the theory's other predictors are measured. For example, "to have predictive validity, the object of [an] attitude must be composed of the same target, action, context, and time elements as the behavior" (F&A 2010, 75).

Outline of This Dissertation

Although there are many philosophical questions we could ask about the RAA, I will focus on three issues: 1) the relationship between attitude and perceived norm, 2) the potential addition of self-identity as a predictor of intention, and 3) the definition and measurement of intention.

In chapter one, I revisit some conceptual issues first raised by Paul Miniard and Joel Cohen over forty years ago. The issues explored in this chapter concern the relationship between attitudes and perceived norms. In agreement with Miniard and Cohen, I argue that insofar as perceived norms motivate action because compliance (or non-compliance) is associated with desirable/undesirable outcomes (broadly construed), the influence of perceived norms is – at least in theory – already accounted for by attitudes, which are overall evaluations based on salient behavioral outcomes. Additionally, again in agreement with Miniard and Cohen, I argue that in counting expert opinions as sources of social pressure, Fishbein and Ajzen fail to distinguish between social pressure and informational influence. This failure is likely to result in unnecessary overlap between attitudes and perceived norms as measured by researchers working within the framework supplied by the TPB/RAA.

In chapter two, I discuss a commonly suggested addition to the TPB/RAA: self-identity. Fishbein and Ajzen argue that many of the items that are used to measure

self-identity appear to measure determinants of behavior that are already included in the theory (e.g. descriptive norm, attitude), and that if this is the case, self-identity should not be added to the theory. I argue that their assessment neglects some of the ways in which items intended to measure self-identity might be interpreted. I contend that the available evidence suggests that self-identity might fit within the theory just as well as perceived norm.

Chapter three focuses on the concept of intention in the RAA. Fishbein and Ajzen define intention as "the subjective probability of performing a behavior" (F&A 2010, 40). In other words, for Fishbein and Ajzen, to intend to do something is to believe you will do it. I argue that 1) this definition does not match our commonsense notion of intention, and 2) behavioral expectations (i.e beliefs about what we will do) are not well suited to fill the role played by intention in the RAA.

Chapter One:

Two Problems in the Reasoned Action Approach

1.1 Introduction

According to Fishbein and Ajzen's reasoned action approach (RAA) to the prediction and explanation of human behavior, attitudes, perceived norms, and perceptions of behavioral control lead to the formation of intentions, which in turn guide behavior (Fishbein and Ajzen 2010, 21). In this chapter, I discuss two problems associated with Fishbein and Ajzen's definitions of attitude and perceived norm. Both of these problems were pointed out by Paul Miniard and Joel Cohen – and too quickly dismissed by Fishbein and Ajzen (hereafter F&A) – over forty years ago (Miniard & Cohen 1979, 1981). In their criticism of the theory of reasoned action, Miniard and Cohen present these problems together as they relate to an overarching concern regarding the inadequate separation between attitude and subjective norm (the predecessor of perceived norm) (Miniard & Cohen 1981, 310). The first problem is that because attitudes are overall evaluations, the behavioral outcome evaluations that motivate compliance with perceived social pressure also contribute to attitudes. This results in some sources of influence being counted twice when the RAA is applied. The second problem has to do with the particular examples of social pressure that Fishbein and Ajzen tend to use when illustrating the concept. One central example – of a doctor influencing a patient – demonstrates a failure to distinguish between social pressure and informational influence. If these problems are to be solved

while the general shape of the RAA is preserved, a new line must be drawn between the normative and attitudinal components of the theory.

1.1.1 Attitudes in the RAA

F&A define attitude as "a latent disposition or tendency to respond with some degree of favorableness or unfavorableness to a psychological object" (F&A 2010, 76). Attitudes are unidimensional; an attitude is represented by a single score along an evaluative continuum (F&A 2010, 77). Disambiguating attitude from affect, F&A write:

> [W]e use the term *attitude* to refer to the *evaluation* of an object, concept, or behavior along a dimension of favor or disfavor, good or bad, like or dislike. Examples of responses reflecting attitude are approval or disapproval of a policy, liking or disliking of a person or group of people, and judgments of any concept on such dimensions as *wise-foolish, enjoyable-unenjoyable, desirable-undesirable, good-bad*, or *pleasant-unpleasant*. (F&A 2010, 78, italics in original)

In the context of the RAA, the object of an attitude is a behavior. Thus, attitudes are based on behavioral beliefs (i.e. beliefs about the likely outcomes of a behavior), weighted by outcome evaluations. F&A explain:

> Our approach to the formation of attitudes relies on the expectancy-value model, which assumes that attitudes are formed spontaneously and inevitably as we form beliefs about an object. In the case of a behavior, these beliefs are mainly concerned with the behavior's likely consequences. (F&A 2010, 126)

In short: an attitude toward a behavior consists of an overall evaluation of the behavior, based on its likely consequences.

1.1.2 Perceived Social Pressure

In the RAA, perceived social pressure – or, interchangeably, *perceived norm* – is "the overall normative influence derived from perceived injunctive and descriptive norms"

(F&A 2010, 133). This concept is intended to capture the influence of a subject's beliefs about 1) what others think they – the subject – should do, and 2) what others do, themselves.

Earlier versions of the theory – the theory of reasoned action (TRA) and the theory of planned behavior (TPB) – included a more limited concept of social pressure, termed *subjective norm*. As it refers to "a specific behavioral prescription or proscription attributed to a generalized social agent" (F&A 2010, 131), the concept of subjective norm captures only the influence of injunctive norms.[3] However, in the RAA – as in Fishbein's Integrative Model (Fishbein, 2000) – the influence of descriptive norms is also included in the normative component, to reflect the fact that we sometimes "experience normative pressure because we believe that important others are themselves performing or not performing [a behavior]" (F&A 2010, 131). Ajzen has explained that the influence of descriptive norms was added to the theory "in recognition of the fact that we can form beliefs as to what is expected of us not only by inferring what important others want us to do (injunctive norms) but also on the basis of the observed or inferred actions of...important social referents" (Ajzen 2012, 17). However, we must be careful not to read too much into this explanation. The fact that Fishbein and Ajzen propose that descriptive norms may be weighted by identification with the referent (F&A 2010, 147) – measured in terms of "wanting to be like" them (F&A 2010, 454)[4] – indicates that perceived social pressure is meant to cover more than just pressure relating to what is

[3] As F&A put it, "[i]njunctive norms refer to perceptions concerning what should or ought to be done with respect to performing a given behavior, whereas descriptive norms refer to perceptions that others are or are not performing the behavior in question" (F&A 2010, 131).

[4] Note that the degree to which one "wants to be like" a social referent is not necessarily the same as the referent's degree of "importance". If someone is an important *negative* social referent, the respondent may wish *not* to be like them.

normatively and empirically "expected of us". It also covers the influence of a person's desire to be like others who are important to them, regardless of whether or not those others have any expectations of them (or even know they exist).

Although it is sometimes called *perceived norm*, perceived social pressure as defined in the RAA does not refer to perceived injunctive and descriptive norms, simpliciter. In the case of injunctive norms, the "pressure" component derives from motivation to comply; in other words, it refers to what might be called *felt* pressure. Similarly, as noted above, Fishbein and Ajzen suggest that descriptive norms might be weighted by the degree to which a person "wants to be like" the referent, although they question the practical utility of this approach (F&A 2010, 148). Practical considerations aside, descriptive norms must be weighted by a motivational factor in order to play the same theoretical role as injunctive norms.

Thus, in order for perceived social pressure to be present, neither the existence of a norm nor the perception thereof is enough. A person is only subject to perceived social pressure insofar as they are motivated to comply with/conform to the relevant injunctive and/or descriptive norm(s). If I believe everyone in the world thinks I should dye my hair pink, but I don't care one bit about what anyone else thinks I should do, and there are no pink-haired people whom I wish to be like (at least with respect to hair color), then – according to the RAA – I am not under pressure at all with respect to this behavior.

1.1.3 Norms, Motivation, and Outcomes

The fact that perceived norms are only influential insofar as they are motivating turns out to be the source of some serious problems.

F&A's theory is an expectancy-value model (Fishbein & Ajzen 2010, 97). However, a brief reflection on the basic assumptions that are central to expectancy-value models reveals that F&A's treatment of perceived norm as a separate, independent determinant of behavior is problematic.

Although individual expectancy-value models differ in their details, the basic premise of expectancy-value theory (EVT) is that motivation to engage in a behavior depends on 1) a person's subjective evaluation of how likely the behavior is to lead to certain outcomes (this is the "expectancy" component), and 2) how much the person (dis-)values those outcomes (this is, of course, the "value" component). There is no reason why this general framework should not apply in cases involving norm compliance. When applied to such cases, EVT states that a person's level of motivation to comply with a norm depends on 1) her beliefs regarding the outcomes that are likely to result from (non-)compliance, and 2) how much she (dis-)values those outcomes. If – all things considered – the expected outcomes associated with compliance are better than those associated with non-compliance, the person will be motivated to comply.

For present purposes, whether or not EVT provides a complete explanation of motivation (to comply with norms or in general)[5] is irrelevant. The argument I wish to make depends only on this point, which I assume to be uncontroversial: in the majority of cases, motivation to comply with norms depends at least in large part on anticipated outcomes, where "outcomes" is broadly construed to include both intrinsic benefits/drawbacks (i.e. those relating to the performance of the behavior itself, e.g. a

[5] There are reasons to doubt that it does. For example, Buchak (2013) argues that expected utility theory (the philosophical sibling of expectancy-value theory) fails to capture the way perceptions of risk affect decision-making.

positive feeling enjoyed while engaging in a favorite exercise) and extrinsic benefits/drawbacks (i.e. those brought about through the performance of the behavior, e.g. improved fitness as a result of exercise) (Studer & Knecht 2016, 28). For example, I may be motivated to do what my mother thinks I should do (i.e. comply with my mother's normative expectation) because I want her to be proud of me (an extrinsic outcome), or because complying with her expectation – which I take to be legitimate – would make me feel good about myself (an intrinsic outcome if I feel good as I perform behavior, and/or an extrinsic outcome if I feel good after the fact).

The relationship between perceived norms and outcomes has not gone unnoticed. While comparing the TRA/TPB to his own social cognitive theory – according to which people are motivated to realize three kinds of outcomes: physical, social, and self-evaluative – Albert Bandura states that both the attitudinal and normative components of the TRA/TPB are operationalized as beliefs about outcomes, explaining that the "attitudinal determinant refers to anticipated costs and benefits of the behavior[, while] the normative determinant refers to anticipated social outcomes" (Bandura 1997, 284).

In response to the idea that normative beliefs are equivalent to outcome expectancies, F&A write:

> It is important to recognize that normative beliefs are quite different from behavioral beliefs or outcome expectancies. 'My having a child in the next 12 months would please my mother' does not represent a normative belief. Instead, it is a behavioral belief (or outcome expectancy) because it associates performing the behavior with a certain outcome (pleasing my mother) and is thus expected to influence the woman's attitude toward having a child in the next 12 months. (F&A 2010, 136)

Normative beliefs are indeed different from outcome expectancies. However, perceived norms as understood within the RAA are not normative beliefs; they are – at least according to the original formulation of the theory – normative beliefs weighted by motivation to comply (F&A 2010, 137). Insofar as motivation to comply with norms depends on the expected/possible[6] outcomes associated with (non-)compliance, Bandura is warranted in claiming that perceived norm, as operationalized within Fishbein and Ajzen's theory, is equivalent to a certain class of outcome expectations.[7]

1.2 The Double Counting Problem

In his analysis of the components of the TRA/TPB, Bandura not only equates perceived norms with anticipated social outcomes; he also equates attitudes with anticipated "costs and benefits". By "costs and benefits", he appears to be referring to physical outcomes, as evidenced by the following passage:

---

[6] By "possible", I mean "subjectively possible". Whether or not the outcome is actually possible is irrelevant.

[7] Illuminating as it is, Bandura's analysis does neglect one nuance of F&A's account. As seen above, Bandura argues that perceived social pressure refers to "anticipated social outcomes" (Bandura 1997, 284). Bandura describes social outcomes as *social reactions*. He writes:

> Human behavior is partly regulated by the social reactions it evokes. Positive and negative social effects form the second major class of outcomes. On the positive side, they include such social reactions of others as expressions of interest, approval, social recognition, monetary compensation, and conferral of status and power; on the negative side, they include disinterest, disapproval, social rejection, censure, deprivation of privileges, and imposed penalties. (Bandura 1997, 22)

However, as F&A point out, social outcomes are not always connected to norms. For example:

> I may believe that buying my wife flowers will please her without believing that she thinks I should buy her flowers. In fact, one reason that buying her flowers may please my wife is that this behavior was neither expected nor prescribed. (F&A 2010, 136)

Perceived social pressure depends only on the evaluation of outcomes that are anticipated to result from compliance with/conformity to injunctive and descriptive norms. It does not refer to anticipated social outcomes in general.

> The attitudinal determinant [of the Ajzen and Fishbein models] refers to anticipated costs and benefits of the behavior; the normative determinant refers to anticipated social outcomes. Social cognitive theory encompasses these two types of outcome expectations plus the influential third type rooted in personal standards and self-sanctions. (Bandura 1997, 284)

Since the three types of outcomes included in Bandura's theory are physical, social, and self-evaluative, we can infer that Bandura takes F&A's "costs and benefits" to be equivalent to his concept of physical outcomes.

If there were such a neat division between perceived norms and attitudes, with one relating exclusively to physical outcomes while the other related exclusively to social outcomes, the two factors might not overlap (assuming physical and social outcomes could be kept distinct). However, attitudes are not based only on the evaluation of anticipated physical outcomes. They are overall evaluations, based on all salient behavioral beliefs (F&A 2010, 98). No class of behavioral beliefs (e.g. those relating to social outcomes) is excluded. Thus, the anticipated outcomes that motivate compliance with norms will also contribute to attitudes. As Miniard and Cohen put it, "any belief that links a behavior to an outcome, regardless of whether the outcome reflects a personal goal or is under the control of others, must be treated as a "behavioral belief" within the Fishbein and Ajzen system and should be represented in the attitudinal component" (Miniard & Cohen 1981, 315).

In their critique of the TRA, Miniard and Cohen argue that the theory "will often be inadequate in separating attitudinal and normative influences" (Miniard and Cohen 1981, 310). They claim that although Fishbein and Ajzen maintain that it is important to distinguish between beliefs about behavioral outcomes and beliefs about the expectations of relevant referents, Fishbein & Ajzen "have offered no evidence that people, in fact,

maintain such a distinction in their thinking" (Miniard and Cohen 1981, 313). Miniard and Cohen illustrate their point using an example originally offered by Fishbein and Ajzen: the case of a parent who believes their child thinks they (the parent) should buy Sugar Puffs. Fishbein and Ajzen regard "my child thinks I should buy Sugar Puffs" as a normative belief, while they categorize "buying Sugar Puffs will please my child" as a (non-normative) behavioral belief (F&A 1975, 304). However, as Miniard and Cohen note:

> [T]he belief, 'Buying Sugar Puffs will please my child' and the belief 'My child thinks I should buy Sugar Puffs,' while structurally different, may reflect a similar underlying concern with the child's reactions. The second belief might result from a direct statement by the child, and the first might very well be a direct implication of that statement. For this reason it is not clear why so much should be made of this surface distinction. What may be much more significant, however, is whether or not these two beliefs reflect a common underlying concern with pleasing the child *for normative reasons* (e.g., to avoid a temper tantrum or have the child respond with praise). (Miniard and Cohen 1981, 313, emphasis in original)

I do not believe that the difference between "Buying Sugar Puffs will please my child" and "My child thinks I should buy Sugar Puffs" is a mere "surface distinction". However, it is true that in terms of their influence on behavior, the two beliefs may function in much the same way, or may represent two stages in a single line of reasoning which results in some degree of influence on behavior (e.g., my belief that my child thinks I should buy the cereal may lead me to believe my child will be pleased if I buy it, and – if I wish to please my child – I may take this as a reason to buy the cereal). If researchers wish to distinguish between normative and non-normative influences on behavior in order to determine the most effective behavior-change strategies, it may indeed be better

to focus on the degree to which normative *reasons* factor into a person's decision-making, rather than on the presence or absence of normative *beliefs*.

However, the problem is not just that the normative component of the theory results in the categorization of behaviors which are not done for normative reasons as normatively influenced. As discussed above, insofar as even truly normative influences on behavior are influential because of their association with certain behavioral outcomes (e.g. the avoidance of a tantrum), they will also influence attitudes. Perceived social pressure derives (at least in part) from the evaluation of anticipated outcomes relating to compliance with injunctive and descriptive norms, while attitude derives from the evaluation of anticipated outcomes in general. Thus, the factors that determine perceived social pressure also influence overall attitude, with the result that perceived social pressure and attitude are not two conceptually independent,[8] non-overlapping determinants of behavior. This is true of attitudes as defined and as measured. As Miniard and Cohen point out, "measures of [attitude] have typically asked subjects to simply evaluate their 'performing behavior X.' Since subjects are not asked for their independent evaluation (i.e., how they feel about the behavior unmindful of normative outcomes), it is likely that subjects will, to some extent, include normative considerations in their responses" (Miniard & Cohen 1981, 316).

Attitudes are not likely to be affected by perceived norms in every case, as it is unlikely that people always consider social/normative outcomes when evaluating behaviors. For example, if the behavior in question is performed in private, social

[8] As discussed in chapter two, Fishbein and Ajzen require that potential additions to the RAA be "conceptually independent of the theory's existing predictors rather than be redundant with them" (F&A 2010, 282).

outcomes may be irrelevant (particularly if social outcomes are conceived of as social *reactions*, as in Bandura's social cognitive theory [Bandura 1997, 22]). However, it is not difficult to imagine behaviors toward which a person's global evaluation (i.e. attitude) would most likely be affected by social/normative factors. For example, my answer to "Would attending your friend's birthday party be good/bad?" may be strongly influenced by anticipated social/normative outcomes (e.g. I may believe that my friend will be pleased with me if I attend, or that I will be failing to meet my friend's legitimate expectations if I do not attend).

Empirical results are compatible with this analysis. O'Keefe points out that attitudes, injunctive norms, and descriptive norms "are generally reasonably positively correlated with each other, with mean correlations ranging from roughly .35 to .60" (O'Keefe 2016, 181), and Ajzen himself notes that "empirical research has typically reported correlations of low to moderate magnitude among [the theory's predictors]" (Ajzen 2020, 319). For example, a recent study of local wine consumption in the Canary Islands found that "[a]ttitude towards local wine consumption [was] directly and positively influenced by subjective norms" (Sabina del Castillo et al. 2021, 12).

In sum: A researcher who successfully measures perceived social pressure and overall attitude as defined within the RAA is likely to capture (at least some of) the influence of perceived social pressure twice. Let's call this the Double Counting Problem.

1.2.1 Rewards and Punishments

F&A acknowledge that some of the beliefs that commonly contribute to perceived social pressure also contribute to attitudes. While discussing the relationship between their

theory and French & Raven's (1959) theory of social power, they point out that "[w]e may comply with perceived social pressure because the social agent exerting the pressure is thought to have the power to reward desired behavior…[or to] mete out punishment for noncompliance" (F&A 2010, 130). Shortly afterwards, they write:

> Cialdini's [...] view implies that in addition to the direct effect of descriptive norms on intentions described earlier, descriptive norms can also have indirect effects. We often have information about the behavior of others that goes beyond simply registering what they are doing. For example, *we may note that their behavior is rewarded or punished by others, and this information can influence attitudes toward the behavior* as well as lead to the inference that the behavior is prescribed or proscribed (injunctive norm). Second, we may learn that the behavior leads to other positive or negative outcomes, again affecting attitudes [...]. (ibid., 132, emphasis added)

In the quoted passages, F&A acknowledge that 1) anticipated rewards and punishments are sometimes responsible for the motivational force of perceived social pressure, and 2) anticipated rewards and punishments influence attitudes.

Fishbein and Ajzen address the fact that rewards and punishments – and other social outcomes – influence attitudes in their 1975 book. They write:

> It can be argued that normative beliefs may be considered a proper part of [attitude]. That is, some of the consequences of performing a given act are that the act may please or displease relevant reference individuals or groups, and that it may lead to reward or punishment from a given referent. Depending on a person's evaluation of these consequences, his attitude toward the behavior should become favorable or unfavorable. The present theory is actually not incompatible with this view. In fact, beliefs of this type may be part of the person's salient beliefs about performing the behavior and thus may influence [...] the person's attitude toward performing the behavior. (F&A 1975, 304)

They go on to argue that "the theory suggests that it is useful to maintain the distinction between beliefs about the consequences of performing a behavior and beliefs about the expectations of normative referents," because, for example, one can hold the belief "my doctor thinks I should take medicine X" without "holding beliefs about the behavior, such

as ‘taking medicine X will please my doctor’ or ‘taking medicine X will lead to a reward from my doctor’” (F&A 1975, 304).

Beliefs about behavioral outcomes are different from beliefs about the expectations of normative referents. The distinction between the two should be maintained. However, that a person can hold a normative belief without holding any particular related behavioral beliefs (or any related behavioral beliefs at all) is beside the point. The problem is that insofar as motivation to comply with norms derives from anticipated outcomes of some kind, the influence of perceived social pressure will be accounted for by its influence on overall attitude[9] toward the behavior. If the influence of perceived social pressure depends *entirely* on anticipated outcomes, the influence of perceived norms is entirely accounted for by the attitude concept, with the result that if the measurement of attitudes were perfectly accurate, perceived social pressure would have no additional, direct influence on intention. In short, the problem is not that normative beliefs are not distinct from attitudes; rather, the problem is that many (perhaps all) of the beliefs that motivate compliance with norms also influence attitudes.

1.2.2 Implications of the Double Counting Problem

Some degree of collinearity among explanatory variables is a “fact of life” (Gujarati 2004, 341). Ajzen notes that although attitude and subjective norm are (in his view) conceptually independent, they are free to – and do – correlate with each other empirically (Ajzen 2020, 319). That being said, the conceptual overlap between attitude

[9] This is true at least of attitudes as they are *defined* by Fishbein & Ajzen, even if it is not true of attitudes as *measured* by researchers working within the RAA, as attitude measures may fail to accurately capture overall evaluations.

and perceived norm described above adds a potential source of collinearity between these two factors, beyond the amount that would be seen if they were conceptually independent. The degree to which the overlap between attitude and perceived norm increases correlation between the two factors as measured will presumably depend on the nature of the behavior under study, the specific measures that are used, and the manner in which subjects interpret and respond to survey items. (For a discussion of the statistical implications of multicollinearity within the context of the TRA, see Miniard & Cohen 1981.)

When evaluating the severity of the Double Counting Problem, it is worth noting that the RAA is intended not only to predict, but to explain human behavior. When it comes to explanation, the consequences of the conceptual overlap between attitude and perceived norm are particularly dire. Unless it is the case that outcomes related to injunctive and descriptive norms are in fact given double weight (being considered once on their own, and again – with added effect – in the context of overall attitudes) by agents engaged in practical reasoning, F&A's contention that perceived norms and attitudes both independently and directly drive the formation of intentions is incorrect.[10]

1.3 The Expert Problem

In the previous sections, we saw that the concepts of perceived social pressure and attitude overlap insofar as motivation to comply with norms derives from the consideration of anticipated outcomes associated with

[10] As mentioned in the introduction, F&A have expressed the "central equation" of their theory as follows: $B \sim I = (A_B)W_1 + (SN)W_2$, where B is the behavior, I is the behavioral intention, A_B is the attitude toward the behavior, SN is the subjective norm, and W_1 and W_2 are empirically derived weights (F&A 1975, 301). However, the equation does not appear in their 2010 book.

(non-)compliance/(non-)conformity. However, the problems with Fishbein and Ajzen's treatment of perceived social pressure do not stop there. In addition to failing to create two truly distinct determinants of behavior by ensuring that the outcomes that drive compliance with norms do not contribute to attitudes as defined within the theory, Fishbein and Ajzen further blur the line between attitudinal and normative influence by using examples of normative influence that are much more plausibly interpreted as examples of informational influence on attitudes. In other words, if we were to try to solve the Double Counting Problem by drawing a line between normative and non-normative influence, some of Fishbein and Ajzen's most central examples of normative influence would belong – for the most part, at least – on the other (non-normative) side of the line.

F&A frequently use the opinion of one's doctor as an example of a source of perceived social pressure. In the previous section, we saw that F&A used a belief about a doctor's opinion as an example of a normative belief in 1975, when their theory was still relatively young. This example remains so central that an individual's doctor appears in the appendix of *Predicting and Changing Behavior: The Reasoned Action Approach* as the sample injunctive normative referent in the standard questionnaire questions presented for reference (F&A 2010, 454). When discussing how a free-response format can be used to elicit a person's salient normative referents, F&A explain that survey respondents might be asked to list 1) the people who would approve of/encourage the behavior in question, 2) the people who would disapprove of/discourage the behavior in question, and 3) the people whom the respondent "might want to talk to" if deciding to perform the behavior in question (F&A 2010, 135). In a hypothetical example involving a

woman's responses to questions regarding the behavior "having a child in the next 12 months", they list the following six referents: husband, priest, mother, best female friend, sister, and doctor (F&A 2010, 136). One of these things is not like the others.[11]

If asked whom they "might want to talk to" before making a decision that could have serious health implications (such as the decision to have a child), it would not be surprising if most survey respondents said they would want to talk to their doctor. If your car starts making funny noises on a deserted backroad a few miles from your destination, you might want to talk with an auto mechanic before deciding whether or not to try to drive it the rest of the way. The reason for this is (presumably) not that mechanics are *salient social referents* – rather, the reason is that auto mechanics are experts about cars, just as doctors are experts about health.[12]

It might seem as though the fact that perceived social pressure involves motivation to comply specifically with doctors' *opinions* about what one *should* do – as opposed to motivation to take the things they say into consideration, more generally – would limit the concept to the realm of what might reasonably be called "social pressure". However, the opinions of doctors are influential primarily because these opinions are taken as sources of information about what would be good for us. In many cases, there is no important difference in motivational force between the statements "I

[11] A person may ask their priest, spouse, friend, or sibling for advice because they believe they will be a reliable source of information about the likely outcomes of a life decision, just as they would ask their doctor for advice about the likely outcomes of a health decision. However, the influence of a doctor *qua doctor* heavily depends on their status as an expert about health. In general, priests, spouses, friends, and siblings tend to have more influence that derives from their social role/relationship to the person to be influenced, although they may also have influence due to their expertise (this may be particularly true in the case of a priest).

[12] As a sample measure of motivation to comply with one's doctor (as an injunctive normative referent), Fishbein & Ajzen offer the item "When it comes to matters of health, I want to do what my doctor thinks I should do", rated on a 1-7 scale from "agree" to "disagree" (F&A 2010, 454).

think you should eat more vegetables" and "eating more vegetables would be good for you", as uttered by a doctor. Although one is presented as an opinion or prescription while the other is presented as a factual statement, they are both likely to be taken as evidence that eating more vegetables would be good for the patient's health.

Miniard and Cohen also argue that Fishbein and Ajzen's treatment of expert opinions is problematic. In their 1981 paper, they point out that the fact that expert opinions are counted as sources of normative influence in Fishbein and Ajzen's model has the potential to result in double counting. They write:

> The Fishbein-Ajzen formulation [...] implies that the influence resulting from the demonstrated expertise of another person, who is valued solely for his/her knowledge on the particular topic, should be reflected in the normative component. But the potential for confounding and double counting exists since these are precisely the factors that should have led to a greater likelihood of information acceptance and hence impact on the attitudinal component. (Miniard & Cohen 1981, 314)

Although the potential for double counting is always present (because, as argued above, even cases of true normative influence involve the weighing of outcomes), the problem is especially salient when the source of (allegedly) normative influence is influential for entirely non-normative reasons.

1.3.1 Normative Influence vs. Informational Influence

It is true that whenever a doctor influences a patient, the doctor is exercising *social power*, understood as the power to exert social influence (Raven 2008, 1). However, the general category of social influence must not be confused with the more specific category of normative influence. Social influence is "a change in the belief, attitude, or behavior of a person (the target of influence), which results from the action of another person (an

influencing agent)" (Raven 2008, 1); it is "any change in an individual's thoughts, feelings, or behaviors caused by other people" (American Psychological Association, n.d.). Any instance of one person influencing another person is an instance of social influence. Normative influence is one kind of social influence; informational influence is another. The kind of influence that flows directly and automatically from (perceived) expertise is informational influence.

The APA Dictionary of Psychology entry gives the following definitions of "informational influence":

> 1. those interpersonal processes that challenge the correctness of an individual's beliefs or the appropriateness of his or her behavior, thereby promoting change. Such influence may occur directly, as a result of communication and persuasion, or indirectly, through exposure to information and comparison of oneself with others[...]. Also called informational social influence.[...]
>
> 2. the degree to which a person's judgments or opinions about an unclear situation are accepted by others as correct. (American Psychological Association, n.d.)

To be regarded as an expert is just to be regarded as someone with a great deal of knowledge or skill in a particular area. Thus – since skills cannot be directly transferred from one person to another – the kind of influence one has *just by virtue of being regarded as an expert* is informational influence.[13] An expert may also have other kinds of influence; for example, "[a] doctor may emphasise his legitimate role as doctor and stress the 'fact' that a patient must obey his doctor" (Raven 1965, 379). Nevertheless, the fact remains that the influence an expert necessarily has *qua expert* is informational influence.

[13] Even those who have purely skill-based expertise (if there is such a thing) can only influence others *qua experts* by acting as a source of information (e.g. by demonstrating that a particular approach to a task is effective, or by trying to verbally describe their actions).

Insofar as the information in question relates to behavioral outcomes, informational influence works by influencing attitudes. Thus, when a norm is taken as a source of information, it can influence attitudes toward the relevant behavior. F&A acknowledge this when they note that descriptive norms can impact behavior indirectly by providing information about behavioral outcomes, thus influencing attitudes (F&A 2010, 132). Like descriptive norms, injunctive norms also influence attitudes when they are taken as sources of information.

It seems clear that F&A intend their concept of perceived social pressure to capture the effects of normative influence, understood as "the personal and interpersonal processes that cause individuals to feel, think, and act in ways that are consistent with social norms, standards, and conventions" (American Psychological Association, n.d.). Perceived social pressure is, after all, the "normative component" of the theory (Ajzen 2012, 17). However, F&A seem to equate normative influence with the influence of injunctive and descriptive norms in general, without regard for the specific ways in which they influence behavior. This is a mistake; as discussed above, the influence of a norm is not always normative influence. Sometimes, it is (non-normative) informational influence. Furthermore, instead of recognizing that a norm can be a source of informational influence even when it is injunctive (e.g. "you should eat vegetables") rather than descriptive (e.g. "everyone eats vegetables") and taking measures to disambiguate between normative and informational influence, F&A exacerbate the problem by using a case of (largely) informational influence – the influence of a doctor's opinion – as a central example of perceived social pressure. I call this problem – the problem of expert opinions being analyzed as sources of normative influence, even when

they are more likely to be sources of (primarily) informational influence – the Expert Problem.

1.3.2 Social Power & Social Pressure

That F&A fail to consistently distinguish between normative and informational social influence becomes apparent upon a careful reading of their discussion of French and Raven's (1959) theory of social power. They write:

> Based on [French and Raven's 1959] analysis it may be suggested that others can influence our behavior because they possess one or more types of power:
>
> 1. *Reward power*: We may comply with perceived social pressure because the social agent exerting the pressure is thought to have the power to reward desired behavior.
> 2. *Coercive power*: Conversely, the social agent may be able to mete out punishment for noncompliance.
> 3. *Legitimate power*: Compliance with perceived social pressure may be based on the belief that the social agent has the right to prescribe behavior due to his or her role or position in a particular group, network, or society at large.
> 4. *Expert power*: We may comply with perceived social pressure because of the social agent's knowledge, expertise, skills, or abilities.
> 5. *Referent power*: Compliance with perceived social pressure may derive from a sense of identification with the social agent, that is, people may comply because they want to be like the agent.
>
> There is an assumption inherent in much theorizing that social norms have no influence on behavior unless they are accompanied by sanctions [...]. In French and Raven's [...] analysis, however, only reward and coercive power involve sanctions to encourage compliance or prevent noncompliance. The other three bases of power produce compliance without using rewards for normative behavior or punishment for violation of social norms. Our reasoned action approach is in agreement with the French and Raven analysis in that we too assume that perceived social pressure can influence behavior even when no rewards or punishments are anticipated. (F&A 2010, 130–31)

These remarks indicate that F&A interpret French and Raven's theory as a theory of the sources of normative influence (which they equate with the influence of social norms). As Miniard and Cohen point out, F&A seem to believe that "anything that increases a referent's social power will increase the likelihood of 'complying' with that individual" (Miniard & Cohen 1981, 315). However, French and Raven's theory is not a theory of normative influence, specifically; it is a theory of social influence in general. Not all cases of social influence involve what might typically be called *social pressure* to comply with norms/directives.

About expert power, French and Raven write:

> The strength of the expert power of [a social agent,] O [over a person,] P varies with the extent of the knowledge or perception which P attributes to O within a given area. Probably P evaluates O's expertness in relation to his own knowledge as well as against an absolute standard. In any case expert power results in primary social influence on P's cognitive structure and probably not on other types of systems. Of course changes in the cognitive structure can change the direction of forces and hence of locomotion, but such a change of behavior is secondary social influence. [...] Accepting an attorney's advice in legal matters is a common example of expert influence; but there are many instances based on much less knowledge, such as the acceptance by a stranger of directions given by a native villager. (French & Raven 1959, 163).

Two things are worth noting:

1) Expert power is the power to influence another person to accept a piece of information. In other words, it is the power to influence another person to update their cognitive structure in particular ways (e.g. form a belief). Any further influence on behavior is secondary.
2) Expert influence is clearly not normative social influence, as illustrated by the example of accepting directions from a stranger.

French and Raven note that "[e]xpert power, where O need not be a member of P's group, is called 'informational power' by Deutsch and Gerard" (French and Raven 1959, 163). Although Deutsch and Gerard do not actually use the term "informational power" in the paper cited by French and Raven, they do discuss "informational social influence", which they define as "an influence to accept information obtained from another as *evidence* about reality" (Deutsch and Gerard 1955, 629, emphasis in original). The purpose of Deutsch and Gerard's (1955) study was to disambiguate between normative and informational social influence and determine the role each played in the experimental results obtained by Sherif, Asch, and Bovard, which were taken to demonstrate the effects of "group" influence despite the fact that "the subjects in these experiments as they made their judgments were *not* functioning as *members* of a group in any simple or obvious manner" (Deutsch and Gerard 1955, 629, emphasis in original). In short: French and Raven (1959) equate expert power with Deutsch and Gerard's informational influence, which – far from being a type of normative influence – is contrasted with normative influence.

Expert influence is not the only kind of informational influence. Although F&A refer only to the five bases of power listed in French and Raven's 1959 paper, Raven's subsequent work includes six bases of power. To the five original bases – reward, coercive, legitimate, expert, and referent – Raven adds informational power. Informational power is potential informational influence, where informational influence is understood as influence that depends on the "content of [a] communication," as opposed to the "nature of the influencing agent" (Raven 1965, 372). In other words, expert power derives from a person's (perceived) status as an expert, while informational

power depends only on a person's ability to convince another person of something through the presentation of compelling statements/arguments.

We have now seen two importantly different uses of the term "informational influence" which we must disambiguate:

1. "Influence to accept information obtained from another as evidence about reality" (Deutsch and Gerard 1955, 629, emphasis removed). This kind of influence includes, for example, influence that depends on what would typically be called "expertise" (e.g. the influence of a doctor's opinion), as well as influence that depends simply on the belief that other people probably have good reasons for acting as they do (e.g. the influence of a descriptive norm when taken as a source of information).
2. Influence that depends on the content of a communication (Raven 1965, 372). This kind of influence does not depend on any beliefs regarding the expertise, knowledge, or good sense of the influencer (i.e. the source of the communication). It depends only on the plausibility of the communicated information itself.

Both expert influence (in French & Raven's sense) and informational influence (in Raven's sense) are kinds of informational influence in Deutsch and Gerard's sense, and as such, both expert power and informational power can be understood as power to exert informational influence (in Deutsch and Gerard's sense) as opposed to normative influence. (Of course, only informational power is the power to exert informational influence in Raven's sense.) Henceforth, I will use the term "informational influence" in Deutsch and Gerard's sense.

The fact that two of Raven's six bases of power (including one of French and Raven's original five) are based in informational – as opposed to normative – influence clearly indicates that French and Raven's theory of social power must not be confused with a theory of the sources of normative influence. That F&A appear to have done so may help to explain their inclusion of expert opinions as sources of normative influence, and their failure to draw a clear line between informational and normative influence in general.

1.4 Evaluating Fishbein & Ajzen's Response

In their "reluctant" response to Miniard and Cohen's concerns (F&A 1981, 340), F&A fail to adequately address the major problems outlined above. As Miniard and Cohen note, although F&A defend the construct validity of their measures, they do not defend the theoretical roles of the constructs themselves (Miniard & Cohen 1981, 333). Although Miniard and Cohen did originally describe their concerns as relating to the validity of the attitudinal and normative constructs of F&A's model (Miniard & Cohen 1979, 102), what is really at issue here is not whether items aimed at measuring perceived norms and attitudes accurately measure what they are intended to measure; what is at issue is, rather, whether or not the theory offers a coherent and useful explanation of behavior.

In defense of the claim that their measures "validly assess two very different constructs" (F&A 1981, 341), F&A write:

> Ajzen and Fishbein (1972) showed that a manipulation designed to influence the attitudinal component had in fact a strong effect on the attitude measure and no significant effect on the normative measure. Similarly, the study showed that a manipulation designed to affect the normative component had a strong impact on

> normative beliefs and only a very slight, albeit significant, effect on attitudes. (F&A 1981, 342)

That attitude and perceived norm are different constructs is clear. Attitudes are overall evaluations of behaviors based on anticipated outcomes, while perceived norms are beliefs about the normative beliefs and behaviors of others. Although they call the difference between them a "surface distinction", I doubt that Miniard and Cohen would argue that these were truly identical. However, as argued in the previous sections, insofar as they influence beliefs about likely behavioral outcomes, perceived norms influence attitudes. Although F&A present the results described above in defense of their theory, they actually lend some support to this claim. If attitudes are overall evaluations which are impacted by behavioral beliefs relating to norms, then perceived norms should have some degree of influence on attitudes, but they should not entirely determine them (as attitudes are also influenced by non-normative factors). Thus, something like the "slight effect" of a norm-related manipulation on attitudes described above is what we might expect. Additionally, since perceived norms are based only on beliefs about the normative expectations and behaviors of others, it would be quite consistent for manipulations aimed at influencing attitudes to have no effect on perceived norms (although they may sometimes have such an effect, e.g. because a behavior that is evaluated more positively seems more likely to be evaluated positively by others).

In narrowly focusing on the uncontroversial claim that behavioral beliefs and normative beliefs are different, F&A fail to address the larger issue. In response to Miniard and Cohen's concerns about expert opinions, F&A point out that it is possible to believe that your doctor thinks you should take a vacation without believing that doing so

will please your doctor (F&A 1981, 343). This is obviously true, but it misses the point. The problem is that insofar as you take your doctor's opinion into consideration because you take it as a source of evidence about likely behavioral outcomes, their opinion influences your attitude, resulting in double counting. Thus, in order for double counting to be avoided in the above example involving the doctor, the doctor's opinion that you should take a vacation would have to be influential in a way that depended on *no beliefs regarding expected behavioral outcomes at all* (including, for example, the belief that taking a vacation would be good for you). That you can hold – and be influenced by – a belief about a doctor's opinion without holding any *particular* belief about behavioral outcomes (e.g. "doing X will please my doctor") is irrelevant.

F&A offer another example in defense of their stance, the psychological implausibility of which might give us pause. They write:

> That behavioral and normative beliefs differ in important ways and reflect more than a "surface" distinction can be seen by considering some additional examples. I may well believe that buying my wife a diamond ring would make her happy, but at the same time also believe that she thinks I should not do so (perhaps because we cannot afford it). In isolation, these two beliefs could produce a positive attitude toward the behavior and a negative subjective norm. (F&A 1981, 343)

Miniard and Cohen argue that this example is "obviously incomplete" (Miniard & Cohen 1981, 336), and I am inclined to agree. They point out that "[the] wife's overall negativity toward his buying the ring stands in marked contrast to [the] belief that such behavior would make her happy" and suggest that it would be more parsimonious to say that the wife's overall judgment of the potential purchase is negative, with her belief that her spouse cannot afford the ring being more heavily weighted than her belief that she would enjoy wearing it (Miniard & Cohen 1981, 336). However, even if we allow that the wife

in this example could be happy to receive a ring she genuinely believed her spouse should not buy for her, it remains the case that her normative belief (that the ring should not be bought) is likely to affect her spouse's behavior primarily or entirely through its relationship with anticipated behavioral outcomes, with the result being that the spouse's positive attitude toward buying the ring will be less positive than it would be if the subjective norm were not negative. This is the key point.

1.5 Moving Forward

If the Double Counting Problem and the Expert Problem are to be solved, a new line must be drawn between the normative and attitudinal components of the RAA, such that informational influence falls clearly on the non-normative side of the line.

In a study investigating the relationship between attitude and subjective norms in which personal outcomes and social outcomes were measured separately,[14] Park found that "when attitudinal components were divided into social attitudes toward a behavior and personal attitudes toward a behavior, only social attitudes were significantly related to subjective norms, while personal attitudes were not" (Park 2000, 171). Miniard and Cohen conducted an experiment in which they replaced the attitudinal component in F&A's theory with a more narrow construct which they called "personal attitude" (Miniard & Cohen 1981, 327). This construct was measured using items such as

[14] Park describes the difference between personal outcomes and social outcomes as follows: "Social outcomes of behavior indicated that the outcomes of behavior had something to do with others, and included getting recognition from others, cutting down on time with friends, making others proud of me, getting good grades that reflect positively on people important to me, and getting a good job that reflects positively on people important to me. Personal outcomes of behavior referred to the behavior that impacted only oneself, not others, including personally learning more, getting stressed, cutting down on time for myself, getting good grades because it is good for me, and getting a job that is good for me." (Park 2000, 167)

"Forgetting what others think you should do and their reaction, you *personally* feel that [performing behavior X is *good–bad*]" (Miniard & Cohen 1981, 322, emphasis in original). Adjusting the attitudinal component of the theory along these lines seems as though it could be a reasonable approach[15] to solving the problems presently under discussion, and the study conducted by Miniard and Cohen provides some empirical support for such an adjustment's ability to reduce the amount of collinearity between attitudes and subjective norms (Miniard & Cohen 1981, 328–29). However, when Miniard and Cohen conducted their experiment, the normative component of F&A's theory included only injunctive norms. The current, expanded state of the theory complicates matters.

As discussed above, the normative component of the theory now includes the influence of descriptive norms as well as injunctive norms. In order to exclude the influence of social pressure relating to both injunctive and descriptive norms, measures of personal attitude might be updated in something like the following way: "Forgetting what others think you should do and their reaction, and disregarding any desire you may have to match the behavior of any people or groups who are important to you, you *personally* feel that...". At this point, though, we might worry that items aimed at measuring personal attitudes are becoming complicated to the point where respondents' willingness or ability to provide accurate responses might be hindered. Evidence suggests that hypothetical

[15] Miniard and Cohen write: "One approach consistent with the expectancy-value form of the attitudinal component might be to view [...] behavioral intentions as being determined by: (I) personal beliefs (i.e., beliefs about outcomes of the behavior that are unrelated to referents' responses) and the values of such outcomes. regardless of the original source of such beliefs and values, and (2) beliefs about the responses of important referents and the value associated with such responses." (Miniard & Cohen 1981, 332)

questions result in more cognitive burden than non-hypothetical questions (Lenzner et al., 2009). Presumably, the more complex the hypothetical, the greater the cognitive burden.[16]

A further problem presents itself. Although the attitudinal component of the theory tends to be more predictive of intentions than the normative component (F&A 1975, 311), perceived norms are sometimes more predictive of behavior than attitudes (for an example, see F&A 1975, 312). This suggests that although attitudes are conceived of as overall evaluations based on salient beliefs about behavioral outcomes, survey respondents may not always consider all kinds of outcomes when responding to measures of general attitude. Assuming that at least some cases in which perceived norms predict intentions/behavior more accurately than attitudes cannot be completely explained by normative influence that is not related to outcomes, it appears to be likely that sometimes – at least to some degree – outcomes relating to perceived norms are left out of the attitude calculation as it is made by survey respondents, with the result that measuring perceived norms separately from overall attitude improves the predictive accuracy of the RAA. It seems unlikely that outcomes relating to perceived injunctive and descriptive norms are the only ones that are ever neglected in this way. Indeed, many potential additions to the theory have been proposed, including moral norms, self-identity, and affective beliefs (Conner & Armitage 1998, 1). As Sniehotta et al. note, in practice, scientists tend to use "extended" versions of F&A's theory, indicating that practitioners do not believe the theory to be acceptable as it stands (Sniehotta et al. 2013, 4). Assuming that at least one of these extensions will turn out to be justified, a solution to the present

[16] If negative social influence (i.e. influence to do the opposite of what others think one should do and/or do, themselves) is regarded as a kind of normative influence, matters will be even further complicated, as – for example – respondents would have to be asked to disregard their desire to fit in with with *or stand out from* "the crowd".

problems that could not accommodate such extensions would be suboptimal. Unfortunately, if personal attitudes are defined as overall evaluations excluding normative influences, then a new overlap problem will be generated if, for example, self-identity is added to the theory (assuming that self-identity also influences behavior at least in part through its association with behavioral outcomes). Excluding new additions to the theory from the personal attitude construct would result in increasingly complex hypotheticals.

An alternative approach would be to leave the attitude concept as it is, while treating perceived norms as a determinant of attitudes (Hale et. al, 270). Smetana and Adler suggested exploring this possibility after finding that the normative component of the Fishbein model had a direct effect on attitudes toward abortion (Smetana & Adler 1980, 95). If the motivational force of norms always depends on their relationship to anticipated outcomes, this may be the "cleanest" approach. Norm-related outcomes are, after all, a subset of all outcomes. However, as noted above, the ability of perceived norms to explain variance in intentions over and above that which is explained by attitudes suggests 1) that the attitude measures that are typically used fail (at least sometimes) to fully capture respondents' overall evaluations of behaviors, and/or 2) that at least with regards to some behaviors in some populations, norms influence intentions *both* directly *and* through their association with anticipated outcomes. If the direct influence of norms on intentions seems likely,[17] this should be acknowledged within the theory. Further additions to the theory could be treated in the same way.

[17] Smetana and Adler found that in addition to affecting attitudes, perceived norms exerted a direct effect on intention. This led them to conclude that "[t]he [behavioral] belief and normative components [...] appear not to be empirically distinct, but neither do they seem to function similarly" (Smetana & Adler 1980, 95).

Conceptually, both approaches are acceptable. Which theory is preferable will depend on the practical utility of the theories' constructs (i.e. the accuracy with which they can be measured and the ways in which they facilitate the creation of effective interventions). Thus, determining the best path forward will require further empirical exploration.

1.5.1 On Defining the Normative Component

While we are considering the separation between the normative and attitudinal components of the RAA, it may be worthwhile to reflect not just on how the normative component should be separated from the attitudinal component, but on exactly what the normative component should include. We might begin by considering F&A's stated justification for including injunctive and descriptive norms. They write:

> It has become clear [...] that normative prescriptions represent only one source of perceived normative pressure. In addition to believing that particular individuals or groups do or do not want us to perform a given behavior, we may also experience normative pressure because we believe that important others are themselves performing or not performing the behavior in question. [...] In contrast to our original use of the term subjective norm, which referred only to injunctive norms, the normative component in [...] our current theoretical framework captures the total social pressure experienced with respect to a given behavior. (F&A 2010, 131)

At this point it is important to distinguish between two senses of the word "normative". Philosophers most often use the term "normative" to mean normative *as opposed to descriptive* (e.g., normative ethics vs. descriptive ethics). This is the prescriptive sense of "normative" – in the above quote, it appears in the term "normative prescriptions". Deutsch and Gerard also use the word in this way when they define *normative social influence* as "an influence to conform with the positive expectations of another" (Deutsch

and Gerard 1955, 629). Some norms (injunctive norms) are "normative" in this sense, while others (descriptive norms) are not. There is another sense of "normative" which is more broad; this is the sense in which "normative" means "conforming to or based on norms" (Merriam-Webster, n.d.), where norms include not only principles and standards, but also "averages", such as "widespread or usual practice[s], procedure[s], or custom[s]" (Merriam-Webster, n.d.).[18] If the word is being used in this sense, "normative pressure" might refer to pressure which is *based on norms* of any kind. Indeed, by "normative pressure", it seems F&A must mean something like "pressure deriving from norms".

Lumping the influence of injunctive and descriptive norms together in this way may seem to make sense insofar as a person's sensitivity to both kinds of norms might depend on the degree to which she feels a need to live up to societal standards. However, as O'Keefe points out, there are reasons to question this approach. One reason is that "injunctive norm[s] and [...] descriptive norm[s] appear to operate in substantively different ways…[f]or example, there is evidence suggesting that whereas the effects of injunctive norms characteristically require some degree of systematic thinking, descriptive norms can operate in ways that require little cognitive effort" (O'Keefe 2016, 195). A second reason is that "[w]hereas there are well established ways of assessing attitude toward the behavior and perceived behavioral control, there are not (yet, anyway) parallel means of assessing generalized perceived norms" (O'Keefe 2016, 196). It is difficult to imagine a straightforward question that would reliably assess a respondent's overall sense of felt pressure to conform with norms of any kind. In practice, in order to

[18] The APA dictionary of psychology offers this definition of "normative": "relating to a norm: pertaining to a particular standard of comparison for a person or group of people, often as determined by cultural ideals regarding behavior, achievements or abilities, and other concerns." (American Psychological Association, n.d.)

calculate perceived social pressure, researchers combine separate measures of injunctive and descriptive norms (e.g. by summing them) (F&A 2010, 183).

As F&A point out, "the most direct test of the utility of a single normative measure is one that compares the predictive validity of a single measure of perceived social pressure with the predictive validity of a model in which intention is regressed on separate measures of injunctive and descriptive norms" (F&A 2010, 182). In defense of their single normative measure, F&A cite a study in which the single measure did "almost as well" as injunctive and descriptive norms considered separately (F&A 2010, 183). If there were indeed "strong a priori theoretical grounds" supporting the single measure (F&A 2010, 182), perhaps "almost as good" would be good enough. However, if the single measure is on shaky theoretical ground (as O'Keefe's arguments suggest), the fact that the single measure didn't do quite as well as the separate measures might be taken as a further point against the single measure.

It is also worth noting that even when considered separately, injunctive and descriptive norms work in a variety of ways. For example, we might comply with an injunctive norm in order to avoid punishment, to make someone we care about happy, to avoid feelings of guilt, or to set an example. There are also many reasons why we might conform to descriptive norms. For example, we may desire to be liked, to avoid making others uncomfortable, or to signal our membership in a particular group. The wide range of possible reasons for compliance/conformity is of particular importance as it relates to interventions. For example, the beliefs/desires that motivate compliance with the normative expectations of close family members are, at least in many cases, probably quite different from the beliefs/desires that motivate compliance with societal/group

norms (as such). Thus, we might expect the interventions that would increase or reduce levels of compliance in each type of situation to be different. For this reason, we might question the value of an approach that lumps the influence of the opinion of a person's mother together with the influence of the "opinion" of society at large (or of whichever segments of society the person considers to be important to them). Perhaps in addition to – or as an alternative to – separating the influence of injunctive norms from the influence of descriptive norms, we should distinguish between what might be termed "levels" of social influence (e.g. interpersonal vs. societal).

I do not wish to advocate for the superiority of any particular definition of the normative. In fact, I suspect that the best normative factor(s) to measure will probably vary, depending on the behavior under study and the purpose of the research. My present aim is simply to point out that although F&A's normative component is maximally inclusive of influences relating to the things we call "norms", it is somewhat arbitrary. If the line between the attitudinal and normative components is to be redrawn, it might be worthwhile to consider whether or not the normative component that most closely matches F&A's current conception of perceived norm is the one that is most likely to produce results that can be used to guide effective interventions.

1.6 Concluding Thoughts

In recent years, some – most notably Sniehotta et al. (2014) – have called for the retirement of the Fishbein-Ajzen model of behavior. The concerns laid out in the preceding sections might be taken as further reasons to retire the theory, rather than as reasons to revise it. However, the TPB remains "one of the most applied theories in the

social and behavioral sciences," having provided the theoretical framework for more than 4,200 published empirical studies as of April 2020 (Bosnjak et al. 2020, 352). Furthermore, a recent meta-analysis showed that TPB-based interventions produced "small to medium" effect sizes (Steinmetz et al. 2016, 225). The authors of that study assert that their results "should lay [the concerns of Sniehotta et al.] to rest" (ibid.). It is possible that minimizing the conceptual overlap between the attitudinal and normative components of the theory will improve the predictive success of the theory and enable practitioners to formulate more effective interventions. It is this possibility that, in my mind, makes ongoing efforts to optimize the theory worthwhile.

With the goal of facilitating more effective interventions in mind, in the next chapter, I will consider the merits of an oft-proposed addition to the theory: self-identity.

Chapter Two:

Self-Identity, Reconsidered

2.1 Self-Identity: A Potential Addition to the RAA

Fishbein and Ajzen's reasoned action approach (RAA) to the prediction and explanation of human behavior is intended to be *sufficient*, in the sense that the inclusion of additional predictors (beyond attitude, perceived norm, and perceived behavioral control) should not improve the theory's ability to predict intentions or behavior (F&A 2010, 281). However, the sufficiency of the theory has often been questioned (Conner & Armitage 1998). Fishbein and Ajzen (F&A) claim that when appropriate measures are used, the theory accounts for ~50-60% of variance in intentions and ~30-40% of variance in behavior (F&A 2010, 282). These numbers may seem low, but F&A argue that due to random error of measurement, the amount of variance contained in well-constructed measures is limited to ~55-65% (F&A 2010, 283). Still, F&A acknowledge that there may be room for improvement in the theory.

F&A are open to additions to the theory that would improve its predictive power. As they point out, the various iterations of the theory have resulted from such additions (the first being perceived behavioral control, and the second being descriptive norm). However, they stress that "for the sake of parsimony, additional predictors should be proposed and added to the theory with caution, and only after careful deliberation and empirical investigation" (Fishbein & Ajzen 2010, 282). The purpose of this chapter is to

contribute some careful deliberation regarding one frequently proposed addition: self-identity.

Because measures of perceived norm tend to account for less variance than might be expected, some theorists have proposed that the normative component of the RAA be extended (F&A 2010, 290). F&A explain:

> Self-identity has been considered as one possible extension [of the normative component]. Relying on ideas from role theory [...] and social identity theory [...], a number of investigators have proposed that people's self-concepts can influence their intentions and actions. The social roles of a mother or father, a student or professor, a nurse or physician normatively entail performance of certain kinds of behaviors. Similarly, social categories can be defined by behavior, such as the categories of smokers, shoppers, or exercisers. Some investigators [...] have argued that people who identify with a certain role or social category are expected to perform, and are more likely to perform, behaviors consistent with that role or category than individuals whose self-concepts do not identify them with the role or category in question. (F&A 2010, 290)

F&A point out that because identification with the relevant group is already considered to affect the importance of perceived norms, "self-identification with a group can serve as a *moderator* of the relation between perceived norm and behavioral intention" within the RAA as it stands (F&A 2010, 290, emphasis in original). However, some theorists have proposed that self-identity be considered a *direct* influence on intention (i.e. a source of influence that is not mediated by subjective norms, attitudes, or perceived behavioral control). (For a review of some work in support of this idea, see Sparks & Guthrie 1998, 1395-1397.) Although F&A admit that empirical results seem to support this view, as adding measures of self-identity to the RAA does seem to improve its predictive power (F&A 2010, 291), they go on to explain why they do not believe current evidence suggests that self-identity should be added to the theory as a direct influence on intentions. In support of their position, F&A argue that many of the survey items used by

researchers do not seem to be true measures of self-identity. In this chapter, I will respond to their assessment of these items and offer reasons to believe that some of them might truly measure aspects of self-identity that are not easily accommodated by the RAA as it stands.

2.1.1 Adding to the RAA: Fishbein and Ajzen's Requirements

Before considering F&A's specific criticisms of self-identity as it has been treated by researchers, it is worth considering the construct's potential fit within the RAA. F&A list five requirements that additions to the RAA should satisfy. Proposed additions should be:

- *Behavior-specific*: "It should be possible to define and measure the proposed factor in terms of the target, action, context, and time elements that describe the behavioral criterion" (F&A 2010, 282).
- *Causally efficacious*: "[I]t should be possible to conceive of the proposed variable as a causal factor determining intention and action" (ibid., 282).
- *Conceptually independent*: "The proposed addition should be conceptually independent of the theory's existing predictors rather than be redundant with them" (ibid, 282).
- *Widely applicable*: "[T]he factor considered should potentially be applicable to a wide range of behaviors studied by social scientists" (F&A 2010, 282).
- *Predictive*: "[T]he proposed variable should consistently improve prediction of intentions or behavior" (ibid, 282).

Let's consider each requirement in turn.

It is theoretically possible to define and measure self-identity in a way that conforms to the principle of compatibility (according to which behaviors should be defined in terms of the target, action, context, and time),[19] thus meeting the requirement of behavior-specificity. However, this approach may yield a variable that is quite removed from self-identity as we normally think of it. For example, it is certainly possible to ask someone if they think of themselves as the kind of person who will "enroll (action) in a continuing education course (target) at a local community college (context) the next time it is offered (time)" (F&A 2010, 44). However, items like this seem highly unlikely to measure something that might plausibly be called self-identity, as opposed to simple behavioral predictions. We intuitively think of our identities as involving the kinds of things we might mention when asked to describe ourselves. As Stryker and Serpe put it, "identities are reflexively applied cognitions in the form of answers to the question '[Who] am I?'" (Stryker & Serpe 1982, 206). Thoits and Virshup explain:

> In essence, social identities are answers to the questions "Who am I?" or "Who are we?" when those answers refer to sociodemographic characteristics (e.g., male, African-American), group/organizational memberships (Little League member, church member), social roles (stepfather, attorney), social types of person (intellectual, leader), and, in some cases, personality or character traits (optimist, caring). (Thoits & Virshup 1997, 107)

While statements like "I am a mother", "I am a Christian", or "I am a runner" are common components of people's self-descriptions, I would venture to guess that nobody has ever spontaneously described themselves as "the kind of person who will enroll in a continuing education course at a local community college the next time it is offered". The traits, interests, roles, and other characteristics we take to be central to our identities tend

[19] As discussed in the introduction, the principle of compatibility is intended to ensure a good match between predictor and behavior, as this increases predictive accuracy.

to be much more general. For example, someone who identifies as a "lifelong student" might be especially likely to enroll in a continuing education course.

Presumably for this reason, studies aimed at investigating the value of self-identity as an additional predictor within the RAA often use measures which fail to satisfy Fishbein and Ajzen's requirement of behavior-specificity. Examples include "I think of myself as a green consumer" (Carfora et al. 2019, 5), "I think of myself as a healthy eater" (Carfora et al. 2016, 26), and "It's in my character to be a smoker" (Hassandra et al. 2011, 1245). However, if self-identity improves the predictive power of the theory despite this failure, this might give us reason to reconsider the principle of compatibility.

The principle of compatibility ensures a good match between predictor (e.g. attitude) and behavior. As Fishbein and Ajzen point out, "[a] change in any one of the elements [target, action, context, or time] constitutes a change in the behavior under consideration"; for example, "buying a dishwasher within a 30-day period is not the same as buying it [within] a 6-month period" (F&A 2010, 30). Attitudes toward buying a dishwasher in a 6-month period – or attitudes toward buying dishwashers, in general – may fail to predict dishwasher purchasing behavior within a 30-day period. In the case of attitudes, the principle of compatibility might be applied without issue, since the respondent is given the opportunity to evaluate the behavior at whatever level of detail the researchers have chosen. Even if the respondent has not considered the behavior at that level of detail before, they are able to do so when responding to the survey. However, it is possible that applying the principle of compatibility to all determinants of behavior – instead of to attitude alone – may do more harm than good. For example, we might find

that in some cases, overly specific definitions of perceived norms might reduce their predictive value. Consider, for instance, a person who forms an intention to go to the doctor within the next month because she believes her mother thinks she should go to the doctor within the next year, and she wants to get it over with. In this case, the perceived norm may be providing most or all of the motivation to engage in the behavior – and yet, if presented with the survey item, "people who are important to me believe I should go to the doctor within the next month", the person might answer in the negative, since she does not believe her mother to have this specific belief. Similarly, someone's self-identity as a runner might motivate them to run twice a week, even if they don't identify as "someone who runs twice a week", specifically. The principle of compatibility was introduced to improve the predictive accuracy of the theory; thus, perhaps it might be reasonable to abandon it in cases where it does not serve this purpose.

Fishbein & Ajzen's second requirement is that proposed additions be causally efficacious. Self-identity seems to meet this requirement as well as any of the existing components of the theory (e.g. perceived norm, perceived behavioral control). We can imagine a person refraining from engaging in some behavior because it would not be consistent with their self-identity, for example. We may not know how mental states cause behavior – or whether they do so at all – but as long as we are assuming that they do, there is no reason to believe that mental states with contents concerning self-identity could not cause identity-compatible behavior.

In chapter one, I argued that the existing components of the RAA are not truly independent, as the anticipated behavioral outcomes that motivate compliance with perceived norms also influence attitudes as defined within the theory. An equivalent

failure of independence applies in the case of self-identity. Just as behavioral outcome evaluations are conceptually distinct from beliefs about the normative expectations and behaviors of others (or such beliefs weighted by motivation to comply), self-identity is conceptually distinct from the theory's existing predictors. However, just as any anticipated behavioral outcomes that motivate compliance with/conformity to norms also influence attitudes (since attitudes are overall evaluations), any anticipated behavioral outcomes that motivate identity-compatible behaviors (as such) are also theoretically accounted for by the attitude concept as F&A define it. Thus, although self-identity is not independent of attitudes as currently defined, its introduction will not create an entirely new problem relating to conceptual independence. The degree of conceptual independence between the components of the RAA is already in need of improvement; if self-identity is added to the theory, conceptual independence should be achieved in a way that accommodates all variables, old and new.

Fishbein and Ajzen's fourth requirement is that additions to the RAA be widely applicable to the kinds of behaviors that are commonly studied by social scientists. Self-identity surely meets this criterion. For example, identities like "runner" or "smoker" may influence health behaviors, while identities like "mother", "man", or "citizen" might influence social behaviors.

Finally, additions to the RAA should be predictive. There is some evidence that self-identity meets this criterion. One meta-analysis of 24 data sets found that "self-identity explained an increment of 6% of the variance in intention after controlling for the TPB components" (Rise, Sheeran, and Hukkelberg 2010, 1). However, Fishbein and Ajzen argue that self-identity as typically defined and measured does not constitute a

genuine addition to the theory. The bulk of this chapter is dedicated to responding to their claims in support of this idea. Specifically, the claims I will be addressing are 1) that some measures of self-identity really measure descriptive norms, 2) that other measures of self-identity really measure aspects of attitude, and 3) that still other measures of self-identity really measure current behavior. Contra F&A, I argue that some of the measures of self-identity that researchers have found to be predictive of behavior are not equivalent to measures of descriptive norms, attitudes, or current behavior. If this is the case, self-identity may be a genuine source of influence on behavior which is not likely to be captured by the theory's existing predictors as currently defined and measured.[20]

2.2 Self-Identity vs. Descriptive Norms

Fishbein and Ajzen believe that some of the items used to measure self-identity are more properly understood as measures of descriptive norm (i.e., as measures that assess the respondent's beliefs regarding the behaviors of other people). For example, they claim that "many investigators would consider 'drinking alcohol is a normal part of everyday life' [...] a measure of descriptive norm" (F&A 2010, 292). I am inclined to agree; "drinking alcohol is a normal part of everyday life" does seem likely to be a report of the perceived descriptive norm, or of current behavior (depending on whose "everyday life" is being reported on), rather than self-identity. Although many of the things we consider to be part of our everyday lives are central to our identities, many are not. For example,

[20] The third predictor in the RAA, perceived behavioral control (PBC), is not mentioned here, as F&A do not argue that measures of self-identity are equivalent to measures of PBC. Furthermore, it is intuitively obvious that self-identity and PBC are distinct.

getting the mail is a “normal part of my everyday life”, but I do not consider this activity to be central to who I am.

Fishbein and Ajzen claim that the other survey items that are typically used to measure self-identity tend, instead, to measure either importance of a behavior (an aspect of attitude) or current behavior. These claims will be addressed later in this chapter. First, however, it will be worthwhile to spend a bit more time on the relationship between self-identity and descriptive norms.

As Fishbein and Ajzen explain, researchers who are interested in self-identity have tended to draw inspiration from role theory (Biddle, 1979; Biddle & Thomas, 1966) and social identity theory (Tajfel, 1974; Turner, 1991). These theories explore the ways that social roles (in role theory) and social groups (in social identity theory) influence our behavior. However, it’s important to note that strong identification with a social type (e.g. intellectual) or trait (e.g. optimism) need not be accompanied by a strong sense of affiliation with the group defined by that type or trait, nor with any desire to conform to group norms, should any exist. As Sparks and Guthrie interpret the term, self-identity refers to “the relatively enduring characteristics that people ascribe to themselves, which take the form of (or incorporate) socially given linguistic categorizations” (Sparks & Guthrie 2006, 1394). According to this interpretation, social identities and role identities are two “particular kinds” of self-identity (Sparks & Guthrie 2006, 1394).

As Fishbein and Ajzen point out, “social categories can be defined by behavior, such as the categories of shoppers, smokers, or exercisers” (F&A 2010, 290). Indeed, these categories do define social groups; it is true that some runners, for example, may feel a sense of community with other runners, and may care about their identities as

runners in part because they value membership in the group. However, it is possible for a person to feel that a behavior is central to their identity without assigning much importance to their membership in the group defined by the behavior. For example, we might imagine a runner who would say, "Running is part of who I am; I don't know what I would do if I couldn't go for a run every morning. When I think of myself as a person, I think of myself as someone who runs, and it's one of the first things I mention when I describe myself to others. I'm not a 'runner' in the stereotypical sense, though…none of my friends are runners, and I don't feel any particular affinity for runners in general. In fact, when I do meet other people who run, I tend to find them annoying." In this case, it would be natural to say that it is the behavior itself that is central to the person's identity, and not the behaviorally-defined social group.

The groups we associate ourselves with are, in many cases, important to our self-concepts. However, a person's self-concept is not – or at least need not be – entirely composed of the various group memberships they conceive of themselves as having. Some aspects of our identities may not only distinguish us from members of other groups, but from other people in general. Consider, for example, a person who considers being the fastest sprinter in the world to be central to their identity, such that they would feel at a loss if they were to lose the title. Certainly, this title may be regarded as a status that is held by a member of the group of runners (or sprinters, specifically), but it need not be thought of primarily in this way. That is, the person might tend to think of themselves not as the fastest member of any particular social group, but rather – quite simply – as the fastest human being (over short distances). Similarly, people who strongly identify as having particular traits – e.g. being shy, honest, or a logical thinker – may or

may not care about being members of the groups defined by those traits. Indeed, some of these "groups" might be better termed "categories"; "honest people", for example, do not seem to form a group in the typical sense.

When a person does not consider an aspect of their identity to be significantly tied to membership in the associated group (e.g. "runners" in the case of running, "honest people" in the case of honesty), measuring the person's degree of identification with the group will not give us insight into the influence of the part of their identity in question. This may partially explain why self-identity seems to have a direct influence on intentions which is not accounted for by perceived norms in combination with degree of identification with the relevant group.

It is also worth noting that, in part because the influence of self-identity is not always related to group membership, it might not make sense to think of self-identity exclusively as an extension of the normative component of the RAA, as opposed to a separate component with both normative and non-normative aspects. To be sure, some aspects of our identity may make us more sensitive to certain injunctive or descriptive norms (e.g. someone who strongly identifies as Canadian might be more inclined to watch hockey than she would be otherwise), and others may relate to internal standards (e.g. someone who strongly identifies as an honest person may take herself to have a moral reason to maintain – and act in accordance with – this aspect of her identity). However, other aspects of a person's identity might not be normative in any standard or straightforward way. For example, a person who defines himself partially through his unique sense of style might not think it would be *bad* if he dressed differently – it just wouldn't be "him".

Perhaps this fact – that we do not necessarily think of the things that are inconsistent with our identity as being *bad* in any general sense – might also explain why, despite the extremely broad definition of attitude, measures of attitude (e.g. those employing differential scales such as *bad–good* or *wise–foolish*) do not seem to fully capture the influence of self-identity. A person who refrains from engaging in a behavior simply because it "wouldn't be *them*" might not consider the behavior to be *bad* or *harmful*, for example. The relationship between self-identity and attitudes will be discussed at length in the next section.

2.3 Self-Identity vs. Attitudes

As discussed in chapter one, attitudes as defined by F&A are overall evaluations of behaviors based on expected consequences. Because no class of consequences is explicitly excluded, attitudes should theoretically be based in part on beliefs regarding likely consequences relating to self-identity (e.g. "If I do this, I will have done something out of character", "If I don't do this, I will fail to fulfill my role as a father"), where applicable. Thus, those who are sympathetic to the basic principles motivating expectancy-value theory might expect the effect of self-identity to be mediated by attitude. However, the fact that including separate measures of self-identity sometimes improves the predictive power of the RAA indicates that the attitude measures that are typically used – e.g. *good–bad*, *pleasant–unpleasant*, *harmful–beneficial* (F&A 2010, 80) – do not fully capture the influence of considerations relating to self-identity. For example, in their study of the effect of self-identification as a "green-consumer" on the consumption of organic vegetables, Sparks and Shepherd found that self-identity had a

direct effect on intentions, contradicting their expectation that the influence of self-identity would be entirely mediated by attitudes (Sparks & Shepherd 1992, 395).

Fishbein and Ajzen believe that some items that are employed to measure self-identity can be regarded as measuring aspects of attitudes which might be captured if an expanded set of attitude measures were used. Specifically, they claim that "importance items" like "blood donation is an important part of who I am" (F&A 2010, 291) "basically assess the extent to which a person values performance of a behavior," and that as such, self-identity when measured in this way "may be conceptually similar to attitude toward the behavior but as an alternative attitude measure may capture aspects of attitude not represented in the traditional semantic differential instrument" (F&A 2010, 292). They go on to point out that if this is the case, then "if importance scales were included in the semantic differential measure of attitude, obtaining a separate measure of self-identity by means of importance items would be of little value." They provide the following example of such a differential scale: "For me to perform the behavior is *important–unimportant, essential–not essential, significant–insignificant*" (F&A 2010, 292, italics in original).

It seems quite likely that including such differential scales as additional measures of attitude would improve the predictive power of the RAA and at least partially capture the way self-identity influences behavior. However, the statement "doing X is an important part of who I am" is significantly different from the statement "for me to do X is important". I consider brushing my teeth to be an important behavior – that is, "for me to brush my teeth is important" – but I don't consider brushing my teeth to be an important part of *who I am*. If we were to discover that everything we thought we knew

about the benefits of tooth-brushing was a fabrication created and propagated by Big Toothpaste, and that in fact, brushing my teeth was bad for my dental health, I would have no problem switching to a new dental hygiene routine (at least as long as my teeth would still feel clean). By contrast, someone who regarded their tooth-brushing routines as *essential to who they were as a person* might be much more distressed. Thus, we might expect the inclusion of differential scales aimed at measuring the importance of a behavior to fail to fully capture the influence on behavior that is measured by survey items that explicitly inquire about a behavior's centrality to a person's self-identity.

If we wished to more fully capture the influence of self-identity by using additional scales to measure attitudes, we may find that it was necessary to use more specific scales, such as "Performing behavior X is [*important to my identity–not important to my identity, consistent with my identity–not consistent with my identity*]". However, such items are too specific to be considered measures of overall attitude. Thus, it appears that measures of attitude as currently defined within the RAA are unlikely to be able to capture the influence on behavior which is captured by measures that inquire into the identity-related importance of a behavior.

2.4 Self-Identity vs. Current Behavior

Fishbein and Ajzen hold that a third category of survey items – including items such as "I think of myself as a cannabis user" – "address the participant's self-concept…by inquiring into the person's actual behavior," and that responses to such items "clearly reflect self-reports of actual behavior" (F&A 2010, 292). Indeed, it is likely that in many cases, actual behavior can be inferred from responses to survey items like this one.

However, the survey items "I think of myself as a cannabis user" and "I use cannabis [often/on a regular basis]" do not necessarily measure the same thing. In some cases, non-users may identify as users. Consider, for example, a person for whom cannabis use is an important part of the spiritual or religious practice they regard as an essential part of their identity, but who currently does not use cannabis due to legal restrictions or lack of access. Such a person may think of themselves as a "cannabis user" who is presently being prevented from using cannabis. Conversely, someone who does use cannabis may not think of themselves as a user. For example, someone who, despite their own occasional use, sees cannabis use in a generally negative light might say "oh, I smoke weed with friends once in a while, but I'm not a real cannabis user". Similarly, we can imagine three people who smoke cigarettes with the same frequency and in the same kinds of situations, but who describe themselves differently: one is an "occasional smoker", one is a "social smoker", and one is a "non-smoker who smokes the odd cigarette with friends".

These considerations demonstrate that survey items which assess the extent to which a person thinks of themselves as someone who engages in a particular behavior do not necessarily measure current behavior. Instead, they are likely – at least in some cases – to measure the extent to which a person *associates* themselves with a particular behavior, or perhaps the extent to which they think of themselves as *the kind of person* who engages in that behavior.[21]

[21] Reluctance to associate oneself with a behavior need not be related to a negative evaluation of (i.e. negative attitude toward) the behavior. In some cases, in fact, such reluctance might relate to a positive evaluation of the behavior in combination with a negative evaluation of oneself. For example, someone who dances every day but has little confidence in their ability might be reluctant to identify themselves as a dancer (i.e. to respond positively to the item, "I think of myself as a dancer"), while an aspiring dancer who has great confidence in their ability might call themselves a dancer despite dancing infrequently, or not at

That being said, it does seem that the more closely a survey item of this kind conforms to the principle of compatibility, the more likely it is to measure current behavior as opposed to self-identity. This is because the more specific an item becomes, the less likely it is to describe a trait that might plausibly be thought of as an aspect of a person's self-identity. Consider, for example, the item "I am the kind of person who gets tested for chlamydia [regularly]" (Booth et al. 2013, 105). It is not clear what "kind of person" this item refers to. Perhaps the relevant kind of person is health-conscious, risk-averse, and/or respectful (of their partners). But if presented with this survey item, is a respondent likely to think of these traits – or any other general traits – and respond as if they had been asked if they possessed them? This seems doubtful. Instead, it seems much more likely that when asked if they are the "kind of person" who gets tested regularly, they would simply base their answer on their actual behavior. According to the simplest interpretation of the phrase, "the kind of person who gets tested regularly" is just someone who *does* get tested regularly. If this is true, the fact that Booth et al. "found that self-identity explained an additional 22% of the variance in chlamydia testing intentions" (over and above the variance explained by attitude, subjective norm, and perceived behavioral control) (Booth et al. 2013, 108) is not surprising.

2.5 Identity Signaling

People do not act in ways they take to be consistent with their identities solely because they wish to maintain their self-concepts and/or positive self-evaluations. Often, people

all. (This effect might be diminished if "dancer" were not a profession, but it is possible to imagine a similar case where "dancer" is replaced with "someone who dances". For example, a person might say "I dance every day, but I still don't think of myself as someone who dances…my dancing doesn't seem good enough to count".)

do things that are (or that they expect to appear to be) consistent with a certain aspect of their identity in part because they want *others* to recognize that aspect of their identity (e.g. "he is a Christian"), or to recognize them as fulfilling it well (e.g. "he is a good Christian"). Thorbjørnsen et al. argue that – at least within the field of communications service marketing research – it may be appropriate to employ an extended version of the theory of planned behavior that focuses on self-identity *expressiveness* (where expressiveness refers to the extent to which a behavior displays the agent's identity and values to others as well as to the agent themselves) as opposed to self-identity per se (Thorbjørnsen et al. 2007, 768). Importantly, a person might wish an aspect of their identity to be recognized by people who neither share nor deeply understand the identity in question. Imagine, for example, a member of religion A who lives in a neighborhood where most people are members of religion B, and who extravagantly decorates their front yard for religious holidays, not because this is something that members of religion A typically do (i.e. decorating in this way is not a descriptive norm for any relevant group), but because they want their religious identity to be seen – they do not want to "blend in" with their neighbors when it comes to this aspect of their identity.

In some cases, people may be more motivated by their desire to be recognized as having – or exemplifying – a certain identity than they are by their desire to actually fulfill the requirements they associate with the identity. For example, someone who wishes to be recognized as a good parent may be more motivated to display the behaviors they associate with that role (giving their child their full attention, for example) in public than they are in private. It is important to note that this desire for recognition may exist in the absence of any related injunctive or descriptive norms. For example, a person may

wish to be recognized as having a certain personality trait (e.g. being "honest to a fault"), even if that trait is neither shared nor admired by the people they wish it to be recognized by, nor by the members of any other relevant reference group.

The implication of this is that two people who identify in the same way may differ in their motivation to act in accordance with their identity in a given set of circumstances. Someone who is highly motivated to maintain their self-concept (i.e. to express their identity to themselves) without regard for social recognition thereof may act in accordance with their identity in private and in public to the same degree, while someone who is highly motivated to maintain their socially-recognized identity (i.e. to express their identity to others) may be more motivated to act in accordance with their identity when they expect their action (or some consequence thereof) to be observed. This is to say that someone who takes some central aspect of their identity (e.g. their religion) to be "a private matter" may be differently motivated than someone who shares the same identity, but considers that identity to be an important part of how they present themselves to – and are interpreted by – others. For example, someone who considers their identity as a blood donor to be a "private matter" may be undeterred by factors that would reduce the visibility of their blood donation (e.g. a remote donation site, the elimination of "I donated blood" stickers), while someone who cared deeply about being seen as a blood donor (even in the absence of any injunctive or descriptive norms favoring blood donation) might be affected by such factors. If this is the case, in addition to measuring self-identity, researchers who aim to produce results that will guide maximally effective interventions may have reason to measure the degree to which people wish their identities to be recognized.

2.5.1 How Should Self-Identity Be Added to the RAA?

I have argued that Fishbein and Ajzen may have been too quick to dismiss the most common measures of self-identity as measures of descriptive norms, attitudes (or aspects thereof), and current behavior. Nevertheless, I believe they are correct in their belief that many of the items typically used to measure self-identity are not appropriate. For example, consider the following items, from a study by Charng, Piliavin, and Pallero (1988) which Fishbein and Ajzen describe as "probably the first study of self-identity in the context of [their] theory" (F&A 2010, 291):

> 1. Blood donation is something I rarely even think about (reversed).
> 2. I would feel a loss if I were forced to give up donating blood.
> 3. I really do not have any clear feelings about blood donation (reversed).
> 4. For me, being a blood donor means more than just donating blood.
> 5. Blood donation is an important part of who I am.
>
> (Charng, Pilliavin & Pallero 1988, 308)

"Blood donation is something I rarely even think about" and "I really do not have a clear feeling about blood donation" are reverse scored, which means that negative responses are taken to indicate that the respondent identifies as a blood donor. However, that a person "thinks about" or "has a clear feeling about" something does not necessarily imply that they personally identify with it, or even that they evaluate it positively. Consider, for example, someone whose very "clear feeling" about blood donation was that they would very much like to donate blood if they were not prevented from doing so (e.g. because of a health condition).[22] Such a person would not be likely to identify as a blood donor, but

[22] This particular reason for a positive answer would be unlikely in the case of the study by Charng et al., as respondents in that study were all people who donated blood at a donation center. Still, the general point stands.

their response to the third item above would be taken to indicate that they did. To see that "thinking about" a behavior need not imply a positive evaluation thereof – let alone personal identification with the behavior – we need only consider how a vegan activist might respond to the survey item, "Meat consumption is something I rarely even think about". Presumably they would respond in the negative, since being a vegan activist generally involves thinking about (other people's) meat consumption. Such a person would also be very likely to respond in the negative if asked if they "did not have a clear feeling" about meat consumption. Sometimes, our clear feelings about things are negative.[23]

If measures of attitude, perceived norms, and current behavior are not likely to fully capture the influence of self-identity, the fact that measures of self-identity tend to increase the predictive power of the RAA gives us reason to believe that adding self-identity to the theory would be beneficial. If this is done, survey items must be worded carefully in order to ensure that self-identity is measured successfully. Only items which respondents would be unlikely to answer in the affirmative (or in the negative, if the item is reverse scored) if the behavior in question were not important to them for reasons relating to their identity/sense of self should be used. For example, surveys might include items like "Blood donation is an important part of who I am", "When I think of myself as a person, I think of myself as a blood donor", or "When describing myself to a new friend, I might say I am a blood donor". Of course, construct validity must be confirmed empirically.

[23] Charng et al. may have felt it was safe to assume that any "thoughts" or "feelings" survey respondents had would be positive, since they were studying repeat behavior among people who donated blood at least once. However, it is generally safer not to make such assumptions. Perhaps some of the donors were attempting to rid themselves of a phobia, for example.

It is important to note, however, that survey items like these are likely to be useful only when an identity is behaviorally defined, or when a behavior is extremely closely associated with a particular identity. Unlike "blood donor" or "runner", most of the classic examples of identities or roles (e.g. man, father, teacher, Buddhist) are not defined in terms of behaviors. Thus, negative answers to measures of self-identity that focus on identification with the specific behavior in question should not be assumed to indicate that self-identity is not a driver of the behavior. For example, a person who identifies as a "masculine man" and believes that a certain conception of masculinity is consistent with this identity might eat meat because he considers it to be a masculine behavior, or – a slightly different thought – he thinks that *refraining* from eating meat is *not* masculine. Such a person's self-identity might cause him to be reluctant to give up meat, even if he wouldn't declare that eating meat was central to his identity. Cases like this are possible – and probably quite common – because engaging in a behavior can be instrumentally important to the maintenance of a certain aspect of a person's identity despite not being in itself constitutive of an aspect of their identity. This is important to understand, as the conditions under which a person's motivations might change are different when a behavior is only instrumentally important to the maintenance of a person's self-identity. If the self-described "masculine" person in the aforementioned example were to change his mind about the masculinity of meat-eating, his behavior might change in the absence of any changes to his self-identity. By contrast, the self-identity of someone who identifies as a "meat eater" will continue to motivate them to eat meat as long as this aspect of their self-identity does not change. Investigations into the influence of self-identity should be designed and interpreted with these considerations in mind.

2.6 Self-Identity vs. Self-Evaluation

We might wonder if the effects of self-identity that are not accounted for by attitudes and perceived norms are accounted for when Bandura's self-evaluative outcomes (Bandura 1997, 22) are taken into consideration. In other words, we might wonder if to suggest that self-identity be added to the RAA is just to suggest that the RAA expand to account for motivation relating to self-evaluative outcomes.[24] This question could be investigated empirically. However, my intuition is that the influence of self-identity is not always strongly tied to self-evaluation, at least as long as such evaluation is taken to involve judging oneself positively or negatively. For example, a person may be motivated to stick with their running routine – instead of switching to swimming, say – because being a runner is part of who they are, even if they don't believe that they would be "worse" if they were a swimmer, or if they switched from running to swimming. It could be argued that insofar as a behavior is rationally guided, there must be something that motivates it – that is, something about the behavior must be evaluated as "good", or at least as good as/better than any alternatives. For example, the aforementioned runner might wish to remain a runner in part because they value having a consistent identity. However, this isn't to say they would evaluate *themselves* negatively if they were to switch from running to swimming, particularly if they had a good reason to make the switch. This point is also illustrated by the fact that it would not be inconsistent for a smoker to say, "Smoking is a part of my identity at this point…I think I'd feel lost if I gave it up. Still, if I wasn't a smoker, I'd probably think more highly of myself". It is possible to become

[24] As discussed in chapter one, Bandura considers his social cognitive theory to encompass the two types of outcome expectancies (physical and social) accounted for – in his view – by F&A's theory, plus a "third type rooted in personal standards and self-sanctions" (Bandura 1997, 284).

attached to an aspect of your identity which you evaluate to be at least slightly negative from a hypothetical unbiased perspective, or from the perspective of an imagined version of yourself who lacks the trait. For example, a person might say "I'm pretty awkward in social situations…it's a flaw, really, but it's part of who I am, so I wouldn't change it." Such a person might enjoy jokes about awkwardness, and may feel a sense of community with other people who consider themselves to be awkward. It would not be nonsensical for such a person to report that they would feel a sense of loss – of self, of community – if they were to wake up and find they had suddenly become socially adept, even if they would not think that their "new self" was worse, or that they had done something bad.

That being said, in many cases, self-identity and self-evaluation are likely to be closely related. Attachment to a trait that is subjectively evaluated as negative is probably relatively rare. Furthermore, many aspects of our identities are accompanied by standards which we would chastise ourselves for failing to uphold. For example, consider two people: both are parents, and both take themselves to have certain duties as parents, but one considers being a parent to be a much more central part of their identity than the other. If the person who more strongly identified as a parent were less likely to fail in their parental duties, this might be explained at least in part by the fact – if it is a fact – that they would evaluate themselves more negatively for failing to uphold these duties. It is conceivable that in cases like this, the effect of self-identity on behavior might be mediated – in part or entirely – by anticipated self-evaluative outcomes.

2.7 Concluding Thoughts

I have argued that the existing components of the RAA are unlikely to fully capture the influence of self-identity on behavior. If this is correct, self-identity may be a good candidate for inclusion in the RAA. However, any such addition should be approached with caution. Measures should be formulated with care, and it must be understood that no matter how it is measured, self-identity is likely to overlap with existing predictors to some extent. Nevertheless, if the addition of self-identity increases the predictive power of the RAA and facilitates the design of more effective interventions, these benefits are likely to make the addition worthwhile.

Chapter Three:

Intentions in the RAA

3.1 Introduction

In this chapter, I consider Fishbein and Ajzen's treatment of intentions as behavioral expectations[25] (i.e. beliefs about one's own future behaviors). As Warshaw and Davis point out, although "intention and expectation regarding future behavior are distinct constructs, easily separable in people's minds", social psychologists – particularly those working within the context of Fishbein and Ajzen's theory – often confound the two (Warshaw & Davis 1985, 214). Fishbein and Ajzen (F&A) believe that this "confounding" is acceptable – that there is no need to distinguish between the two concepts within the context of their Reasoned Action Approach (RAA). My aim is to show that this assessment is incorrect.

The chapter will proceed as follows. The first part of the chapter concerns the relationship between intentions and behavioral expectations. In section 3.2, I argue that what might be called commonsense intentions – i.e. intentions as conceived of within our "commonsense psychological framework" (Bratman 1984, 375) – are not behavioral expectations. I offer two arguments in support of this claim. First, I argue that because

[25] In what follows, wherever I refer to a behavioral expectation in relation to an intention, I am referring to the *corresponding* behavioral expectation, i.e. the expectation that you will do the intended action. So when, for example, I say "an intention is not identical to a behavioral expectation", I mean "an intention to X is not identical to a behavioral expectation that you will X". That an intention to X is a behavioral expectation that you will Y, where Y is not identical to X, is not a view that – as far as I know – anyone holds.

behavioral expectations do not imply corresponding intentions, intentions cannot be *mere* behavioral expectations. Second, I argue that because strength of intention is not identical to strength of behavioral expectation, intentions cannot be behavioral expectations of any kind. The second half of the chapter concerns the role of intentions within the RAA. In section 3.3, I give reasons to believe that "commonsense intentions" are better suited than behavioral expectations to fill the role that intentions are meant to play within the RAA. In section 3.4, I discuss the suitability of various measures that are often used to assess intentions.

3.1.1 F&A's Definition(s) of Intention

In *Predicting and Changing Behavior: The Reasoned Action Approach*, F&A offer two definitions of intention. First, they write: "our use of the term *intention* in this book [...] refers to the subjective probability of performing a behavior" (F&A 2010, 40, italics in original). Because they state that this subjective probability is the "essential underlying dimension characterizing an intention" (F&A 2010, 39), I will call this their *primary definition* of intention. They also offer the following definition, which I will call their *expanded definition*: "in this book we use the term *intention* to refer to readiness to engage in a behavior, a construct that incorporates such concepts as willingness, behavioral expectation, and trying" (F&A 2010, 43, italics in original).

I do not think it is uncharitable to say that F&A offer two definitions of intention. After all, in both cases, F&A claim to be specifying the way in which they will use the term in the relevant work ("in this book…"), and in both cases, the term being defined is italicized. Their use of two different definitions is in itself concerning. Still, the use of

two definitions might not cause serious problems if both obviously referred to the same thing. However, that F&A's primary and expanded definitions of intention refer to the same thing is not obvious. In fact, it is difficult to see how they *could* refer to the same thing, as one identifies intention with behavioral expectation itself (i.e. the subjective probability of performing a behavior), while the other identifies intention with a construct that *involves* behavioral expectation *among other things* (willingness and trying).

Although F&A's two definitions of intention are not obviously compatible with each other, they do both center on the notion of behavioral expectation. Expanded definition notwithstanding, F&A claim to "define intentions in terms of a subjective probability dimension" (F&A 2010, 39). Thus, in what follows, I will focus on their primary definition of intention, according to which an intention is a behavioral expectation. First, I will argue that this definition does not match our commonsense notion of intention. A commonsense intention is not a behavioral expectation. Second, I will argue that it is commonsense intentions, and not behavioral expectations, that are best suited to fill the role of "intention" in the RAA.

3.2 Intentions vs. Behavioral Expectations

The question of how intentions are related to behavioral expectations has divided philosophers. Some think that intending to ϕ implies believing you will ϕ (e.g., Harman 1997); some think that it implies only believing that your ϕing is possible (e.g., Wallace 2001, 20). A philosopher who posits an especially close tie between intention and behavioral expectation is David Velleman. Velleman argues that an intention is a behavioral expectation; he holds that to form an intention is to predict your own behavior,

or in his words, "[an] agent's intention to act, a practical conclusion, [...] consist[s] in a factual conclusion – namely, the agent's anticipation of his next action" (Velleman 1985, 52). However, it is important to note that for Velleman, intentions form a *particular class* of behavioral expectations. Specifically, intentions are "self-fulfilling expectations that are motivated by a desire for their fulfillment and that represent themselves as such" (Velleman 1989, 109). Thus, not *all* behavioral expectations are intentions, on Velleman's view. By contrast, according to F&A's definition of intention, intentions are behavioral expectations, quite simply. No particular class of behavioral expectations is specified or excluded. While Velleman's more nuanced view may not be vulnerable to some of the following objections, I believe they present serious problems for F&A's definition of intentions as what I will call *mere* – or simple – behavioral expectations.

3.2.1 Behavioral Expectations Do Not Imply Intentions

If intentions were mere behavioral expectations, then to expect to do something would be to intend to do it. Thus, a behavioral expectation (e.g. "I expect to go to the store tomorrow") would imply a matching intention (e.g. "I intend to go to the store tomorrow"), as the two would be the very same thing. However, it is easy to see that behavioral expectations do not imply intentions in this way. For example, consider a person who has noticed that, despite her best efforts, she often says things that upset a particular friend of hers. If she were planning on having dinner with that friend, she might say, "I expect to say something that upsets my friend at some point this evening, even though I don't intend to."

For those who find the above example unconvincing because the action is described in terms of its outcome, here is another example: a young person who is asked if they expect to eventually try to have a child may say that they believe they will, even if they have not yet formed any intentions regarding having children. They may have this belief because at this point, the best evidence they have regarding their own future behavior is just other people's behavior. To infer intentions from a child's predictions regarding their future behavior would be a mistake in many cases, and the same is true of people at any age. We can predict our own actions without forming any intentions regarding those actions. Thus, behavioral expectations do not imply intentions. This implies that intentions are not mere behavioral expectations.

3.2.2 Do Intentions Imply Behavioral Expectations?

There is also reason to doubt that intention implies behavioral expectation. Consider Mele's example:

> Al is a forty-five percent shooter from the foul line, and he knows this. He sets himself for a free throw, believing that there is a less than even chance that he will make the basket. The attempt is successful. Surely, barring causal deviance, the great majority of English speakers would count Al's making the basket as an intentional action. (Mele 1989, 19–20)

Here, Al's intention is not accompanied by a belief that he will make the basket. If we ask him if he expects to make the basket, he will deny that he does, as he realizes the odds are against it. He will, however, say that he expects to *try* to make the basket. Perhaps, then, an objector might argue that his intention is just to try, and that this intention does imply a belief that he *will* try. But what will we say if Al does make the basket? Will we say that this was not – strictly speaking – his intention, as his intention was merely to try to make

the shot? If, on the other hand, he misses, will we say that he has achieved what he intended, as all he intended was to try to make the basket? Surely Al does intend to try – his trying is not unintentional – but he also intends to perform the action he is trying to perform.

This example is vulnerable to at least two kinds of objections. Readers who doubt that we can intend outcomes (e.g. making the basket) might object, claiming that Al can only intend his own bodily action (aiming at the basket, releasing the ball, etc.), which he does believe he will carry out. Others might object that although we *can* intend outcomes, the example depends on an illegitimate conflation of goal intentions (which "[focus on] achieving desired goals") and behavioral intentions (which "focus on engaging in a behavior or action") (Conner & Norman 2022, 1). Perhaps although you can intend to achieve an unlikely (but not impossible) goal (e.g. Al can intend to make the basket), you cannot intend to engage in a behavior you do not expect to engage in (so while Al can intend to make the basket, this is not an example of a *behavioral* intention).

In response to these concerns, it is possible to construct a similar case in which the improbable action is the bodily movement itself. Consider, for instance, Setiya's example involving a person who intentionally clenches their fist while in the process of emerging from paralysis (Setiya 2008, 390). This person's doubts regarding their own ability to clench their fist are so strong that – since they still cannot feel their own hand – they do not believe they are clenching their fist even as they are in fact doing it.[26]

[26] Setiya, being himself a cognitivist about intention, offers the following solution to the problem this example seems to pose for those who hold that intention implies belief: "If A is doing ϕ intentionally, A believes that he is doing it or is more confident of this than he would otherwise be, or else he is doing ϕ by doing other things for which that condition holds." (Setiya 2008, 391) To extend this to behavioral intentions (i.e. intentions for the future), we might say "to intend to ϕ is to believe your doing it is more likely than it would be if – other things being equal – you did not intend to ϕ". That intending to do

However, this is no reason to say the clenching is unintentional. If an intentional action is always the result of a prior intention, the fact that you can intentionally do something you did not expect to do implies that you can intend to do something you do not expect to do. If this is the case, intention does not imply expectation. However, this argument will not convince those who do not believe that intentional actions are always preceded by behavioral intentions (i.e. intentions for the future). If intentional actions are not always preceded by behavioral intentions, then the fact that you can intentionally clench your fist even if you don't believe you're doing so does not imply that you can *intend to* clench your fist if you don't believe you *will* do so.

3.2.3 Strength of Intention Is Not Strength of Expectation

That behavioral expectation does not imply intention is clear. This shows that intentions are not *mere* behavioral expectations. However, we have not entirely ruled out the possibility that intention implies behavioral expectation. Thus, we have not yet ruled out the possibility that although not all behavioral expectations are intentions, all intentions are behavioral expectations. In other words, the arguments that have been made this far are compatible with the idea that intentions are a *class* of behavioral expectations. If intentions are a class of behavioral expectations, perhaps F&A's definition is not too far off the mark. However, I believe that a brief consideration of intention strength will cast some doubt on the idea that an intention is a kind of behavioral expectation.

---

something implies believing that your doing it is more likely than it would be had you not formed the intention (holding all other relevant facts static) seems plausible enough. However, this belief does not amount to a behavioral expectation (i.e. a belief that you *will* do it, or that your doing it is *more likely than not*).

Let's consider the following case. Joy lives in a small town where there is just one restaurant: the Middling Mushroom, or the Mush for short. She doesn't feel like cooking this evening, so she forms an intention to go to the Mush. She's not very keen on their food; if she had any other option, she'd eat somewhere else. However, since the Mush is the only restaurant in town and she has no reason to believe that anyone else is going to offer to cook for her, she is almost certain that she will go to the Mush. In other words, she has a very strong expectation that she will go. Luckily for Joy, just as she is preparing to drive to the restaurant, she is surprised to get a call from a friend whom she had believed to be out of town. The friend invites her over for dinner, and she eagerly accepts almost before the friend has finished extending the invitation. Has Joy abandoned a strong intention to eat at the Mush? It seems not. As evidenced by the fact that it was so easily abandoned, her intention was not strong. However, her behavioral expectation *was* strong, as she firmly believed that she would eat at the restaurant.

If this interpretation of the example is correct, the strength of Joy's intention is not identical to the strength of her behavioral expectation. This implies that Joy's intention to go to the restaurant is not identical to her expectation that she will go, as one has the property of being strong while the other has the property of being weak. That the strength of an intention need not match the strength of the related behavioral expectation implies that an intention to ϕ is not a behavioral expectation that you will ϕ.

3.2.4 F&A's Defense of Intentions as Behavioral Expectations

F&A do consider the possibility that there may be a distinction between intention and behavioral expectation, noting the work of Warshaw & Davis (1985). In response to the

suggestion that these concepts should be kept distinct, Fishbein and Ajzen state that "the utility of such distinctions for the prediction of behavior is ultimately an empirical question", and that "available evidence to date suggests that there is little to be gained by the proposed distinctions" (F&A 2010, 41).

Regarding the distinction between intentions and behavioral expectations, Fishbein and Ajzen explain that "[i]t was hypothesized [by Sheppard, Hartwick, & Warshaw (1988) and Warshaw & Davis (1985)] that behavioral expectations are better predictors of behavior than behavioral intentions because the former are more likely to take into account possible impediments to performance of the behavior", and point out that although an initial meta-analysis (Sheppard et. al, 1998) supported this hypothesis, "[s]ubsequent research...has often reported little difference, or even a difference in the opposite direction" (F&A 2010, 41). They take this to support the idea that there is no need to distinguish between intentions and behavioral expectations.

This reasoning seems flawed for two reasons. First, assuming the arguments in sections 3.2.1–3.2.3 are sound, it is quite simply not possible for behavioral expectations and intentions to be the same thing. Given that the RAA is an explanatory theory – and not just a predictive one – the fact that two things are not the same is reason enough to distinguish between them. Second, the very results that F&A reference as supporting their approach may be interpreted as further reason to maintain a distinction between behavioral expectations and intentions. F&A seem to think that the fact that sometimes behavioral expectations and intentions are equally good predictors of behavior, sometimes behavioral expectations are better predictors than intentions, and sometimes intentions are better predictors than behavioral expectations gives us reason to think these

things are equivalent, at least for present purposes. However, one might think, instead, that the following is an interesting question: Under what circumstances are intentions more reliable predictors of behavior than behavioral expectations (and vice versa)? If intentions are not distinguished from behavioral expectations, this question cannot be investigated.

3.2.5 Measures of Behavioral Expectation Should Not Be Used to Measure Intentions

As far as I know, nobody who has given serious philosophical consideration to the matter has defended the view that commonsense intentions are mere behavioral expectations. However, a considerable number of people subscribe to Velleman's much more nuanced and well-developed view, according to which intentions form a *class* of behavioral expectations. Although I believe the above argument regarding intention strength gives us reason to think that this view is mistaken, I do not take myself to have definitively shown that Velleman's view – or any other view according to which intentions are a specific kind of behavioral expectation – is false. However, I do believe that even if such a view is true, it is unlikely to have significant implications for the way intentions should be measured within the RAA framework. Even if intentions are in fact self-fulfilling beliefs about future behaviors that represent themselves as such, we are not likely to be able to base useful measures on this definition, or on any similarly complex or unobvious definition. Surely, the survey item "I intend to go to the store" will do a much better job of assessing commonsense intentions than the item, "I have a self-fulfilling belief that I will go to the store", or "I have a belief that I will go to the store that I expect to be self-fulfilling".

At this point, I hope to have established the following points. First, *mere* behavioral expectations are not – and do not imply – intentions. Therefore, the expectation-based measures of intention that are commonly used by researchers working within the RAA framework (e.g. "I will go to the store") do not measure intentions. Second, even if intentions are a *kind* of behavioral expectation, it is nevertheless the case that the most reliable measures of intention will be those that inquire about intentions as commonly understood, rather than those that are based on an unfamiliar (to most people), complex, and/or controversial philosophical definition. This is to say that if researchers wish to measure intentions, they should use measures that are clearly and closely related to the commonsense notion thereof. In section 3.4, I discuss what these measures might look like.

A further point must be noted. Although I have argued that intention is not mere behavioral expectation, I have not shown that it is intention – and not behavioral expectation – that is best suited to fill the role that is currently played by "intention" in the RAA. I address this issue in the next section.

3.3 The Role of Intention in the RAA

Although F&A define intention in terms of behavioral expectation, the fact that the construct itself is called "intention" – and not "behavioral expectation" – is telling. It appears that F&A mean their theory to fit with the widely held notion that intentions cause behavior. This is to say that in defining intention as behavioral expectation, they seem to be proposing a definition that is meant to match the thing we generally call "intention", as opposed to specifying a technical definition of "intention" that need not

have any relationship to the concept of intention as it is typically understood. Although F&A seem to be somewhat committed to the idea that intention is – or at least centrally involves – behavioral expectation, I suspect that if they had to choose between defining intention differently and removing the notion of intention from their theory altogether, they would choose the former. Still, they do define intention as behavioral expectation. Thus, the possibility that it is this notion – and not the commonsense notion of intention – that should remain in the theory should be considered.

There are at least two reasons to reject this approach. The first is that to remove intentions (by which I mean "commonsense intentions", i.e. the mental states we take ourselves to be reporting when we claim to have intentions) from the RAA – which is supposed to offer a complete explanation of behavior – would (in the absence of further adjustments) be to claim that they have no role in producing behavior. To hold that intentions do not cause the corresponding behaviors – i.e. that there are no "effective intentions" (Mele 2009, vii) – is to subscribe to what Mele calls "scientific epiphenomenalism" about intentions (Mele 2018, 3). According to the scientific epiphenomenalist about intentions, neither conscious intentions nor their physical correlates cause the corresponding behaviors (loc. cit.). This must not be confused with philosophical epiphenomenalism, according to which conscious mental states (as opposed to their physical correlates) are entirely causally inefficacious (loc. cit.). The belief that intentions do not cause the corresponding behaviors is compatible with the belief that they cause other things, such as our verbal reports regarding our intentions (Mele 2018, 2). Thus, scientific epiphenomenalism is more bold than philosophical epiphenomenalism in one sense (as it involves the claim that even the physical correlates of conscious

intentions do not cause the correlated behaviors), but more restrained in another (as it does not imply that conscious intentions are entirely causally inert).

As is well known, the results of experiments conducted by Benjamin Libet and his colleagues in the early 1980s have often been taken to imply that conscious intentions do not cause behavior. However, many have argued that the results do not in fact have this implication (e.g., Flanagan 1992, 136–138; Zhu 2003; Mele 2009). Furthermore, suppose it followed that, if the neural process that results in voluntary behavior begins before the agent is aware of a conscious intention (as Libet's experiments seem to show), then intentions do not cause behavior. It seems likely that if this were the case, similar experiments could produce results that would imply that behavioral expectations did not cause behavior. It would be surprising if participants in a Libet-style experiment reported *expecting* to engage in a behavior (at a particular time) before intending to do so. But if behavioral expectations do not arise earlier than intentions, then if we had reason to take Libet's results to mean that intentions do not cause behavior, we would also have reason to think that behavioral expectations do not cause behavior.

The idea that intentions cause correlated behaviors has great intuitive appeal. Arguments according to which this intuition is mistaken are largely inspired by Libet's experiments; however, the same style of argument could (probably) be made against the idea that behavioral expectations cause behavior. Thus, our intuitions give us reason to think that if a conscious mental state is to play the role that is presently played by "intention" in the RAA, intention itself – and not behavioral expectation – is the best candidate.

The second reason to avoid an approach that replaces intention with behavioral expectation is that it is difficult to see how behavioral expectations could play the appropriate causal role.

Those who subscribe to Velleman's theory of intention do believe that behavioral expectations (of a particular kind) cause the correlated behaviors. As mentioned above, for Velleman, intentions are self-fulfilling behavioral expectations. However, the causal story Velleman tells relies on the claim that our desire for self-knowledge motivates us to act in accordance with our self-predictions. Yaffe helpfully summarizes Velleman's position on this matter as follows:

> Under Velleman's theory, we all have a desire for self-knowledge which is satisfied (although not exhausted) when it turns out that a belief which we have about our current or future conduct is true. So, if I believe that I will go bowling, and then I do, in acting I satisfy my desire for self-knowledge by turning a mere belief about myself into knowledge. Therefore, expectations about what I will do – that is, beliefs of the form 'I will *A*' – provide reason for action, for by acting as I expect I will satisfy my desire for self-knowledge. (Yaffe 1995, 108)

This view is not without its weaknesses. If my desire for self-knowledge gives me reason to do what I expect to do, this implies that my belief that I will do something I do not wish to do gives me a reason to do it. Consider, for example, Bratman's "pessimistic actor":

> A pessimistic actor [...] might be on stage with the purportedly self-fulfilling belief that [A] I will stumble on my next lines as a result of this very belief. This belief that [A] puts the pessimistic actor "into a position to increase [his] self-knowledge by satisfying that very belief." After all, if he does stumble on his lines as a result of his belief that [A], this will be as he expects. (Bratman 1991, 126)

Bratman offers this example in support of the claim that Velleman's view problematically implies that some mental states that are not intentions give us reason to act in just the

same way that intentions do (this is why the fact that the actor takes his belief to be self-fulfilling is important). However, for present purposes, the key point is just that Velleman's view implies that in general, our beliefs about our own future behavior give us reason to behave as we expect (regardless of whether or not they give us reasons in exactly the same way as intentions) (Yaffe 1995, 110–11). The view does not imply that our beliefs about our own behavior always give us *overriding* reasons to behave as we expect; it is possible, for example, that although the actor's desire for self-knowledge gives him a reason to stumble on his lines (because in doing so, he will make it the case that his belief was an instance of knowledge), he has stronger reasons to avoid stumbling on his lines (if possible). Still, that the actor's pessimistic belief gives him any reason at all to stumble on his lines seems doubtful.

In general, when we say we wish to have knowledge – or to have true beliefs – what we mean is that (perhaps among other things) we want our beliefs to match the state of the world. Beliefs have a "mind-to-world" direction of fit. If a belief turns out to be false, our inclination is to update the belief, rather than to change the world. The desire to have knowledge – in a particular domain or in general – is not typically thought of as a desire that the world match one's beliefs, whatever they may be; rather, it is a desire that one's beliefs match the world as it is. For example, if I desire to have true beliefs – or knowledge – about climate change, this does not imply that I want my most terrible predictions to come true, no matter how firmly I believe that they will. For the desire for self-knowledge to be different from the desire for any other kind of knowledge in this respect would be surprising. At least on the face of things, it seems quite possible for a

person who says "I believe I will fail to keep my resolution, but I wholeheartedly hope I am wrong about this" to be telling the truth.

Any theorist who claims that behavior is caused by behavioral expectations is on the hook to explain how this could be the case. Perhaps such an explanation can be provided. However, the primary candidate – an explanation according to which our self-predictions give us reason to bring about their truth through our actions because we desire to have true beliefs (about ourselves, or perhaps in general) – appears to be on shaky ground. In the absence of a positive reason to think that behavioral expectations are more likely to cause behavior than intentions (we are assuming now that intentions themselves are not behavioral expectations), the difficulties associated with the idea that behavioral expectations cause behavior seem to give us reason to prefer a theory that does not include the claim that behavioral expectations – and not intentions – are the immediate antecedents of behavior.

3.4 Proposed Definition and Measures

I have argued that the concept of intention that is included in the RAA should match our commonsense notion of intention; that is, intentions as conceived of within the RAA should be the things we claim to have when we say we have intentions. The precise nature of those things – the things we refer to as "intentions" – is the subject of ongoing philosophical debate. However, there is no reason for the RAA to involve a commitment to any particular philosophical theory of intention. Thus, my present aim is not to present a philosophical definition of intention; rather, it is to present an intuitively plausible definition of commonsense intention that does not imply commitment to any particular

philosophical theory. This is to say that I mean to present a definition of the form "to intend is to X" which is open to further elaboration, such that the holder of a particular philosophical theory of intention could intelligibly say, "yes, to intend is to X, and to X is to…". Perhaps it might be more accurate to say that my aim is to present a guiding notion of intention, as opposed to a strict definition.

As a starting point, let's consider the definition that Warshaw & Davis use in their (1985) paper aimed at "disentangling" intention and behavioral expectation. Drawing inspiration from the definition of intention presented in Webster's Dictionary,[27] Warshaw and Davis define intention as "the degree to which a person has formulated conscious plans to perform or not perform some specified future behavior" (Warshaw & Davis 1985, 214). Although this definition appears to be on the right track, it seems to involve a category mistake, as it identifies an intention (an object) with a "degree" (a property). Perhaps the following definition might be better: an intention is a *conscious plan to perform or not perform some future specified behavior*, where intention *strength* is the *degree of commitment* to such a plan.

A potential weakness of this definition is that it involves the claim that an intention is a plan. This definition may seem to favor Bratman's (1987) view of intentions as plans over, for example, Davidson's (1978) view of intentions as all-out evaluative judgments. Additionally, depending on how one interprets the phrase "conscious plan", it may seem to rule out cases in which a person forms an intention to do something (i.e. she "plans to do it") before formulating a specific plan regarding *how* she will do it. A more open-ended definition would be preferable.

[27] Intend: "to have in mind as a purpose; plan to do, use, give, etc." (quoted in Warshaw & Davis 1985, 214).

Perhaps the following definition might be more suitable for our purposes. An intention is a *conscious (or consciously accessible)*[28] *decision to perform or not perform some future specified behavior*, where intention strength is degree of commitment to such a decision. This definition seems to fit reasonably well with Merriam-Webster's current definition of intention as "a determination to act in a certain way" (Merriam-Webster, 'intention'). Furthermore, this definition leaves open the possibility that intentions might be properly analyzed as plans (if to decide to ϕ is to plan to ϕ) or judgments (if to decide to ϕ is to judge that to ϕ would be good/best). It is even compatible with a cognitivist view according to which to intend to do something is to believe you will do it (if to decide to ϕ is to form a belief that you will ϕ), or to believe that your doing it is more likely than it would be had you not formed that intention (if to decide to ϕ is to increase your credence in the proposition that you will ϕ). For the reasons outlined in section 3.2, I believe that a cognitivist view of the former kind is unlikely to be correct. Nevertheless, that the proposed definition (or "guiding notion") of intention is open to such a view without implying it is, I think, a virtue.

Let us assume, then, that to intend to do something is to have decided – at least tentatively[29] – to do it. Our next task is to determine which measures can be used to assess intentions thus understood. In the following section, I consider the various measures of intention F&A suggest researchers use.

[28] I am assuming that 1) intentions are consciously formed, and 2) you can continue to have a previously formed intention even when you are not consciously thinking of it.
[29] That the decision can be tentative is important. Although the statement "I intend to do it, but I have not fully decided to do it" may be perfectly intelligible, "I intend to do it, but I have not even tentatively decided to do it" seems to involve a contradiction.

3.4.1 Measuring Intentions

I have argued that it is commonsense intention – and not behavioral expectation – that is best suited to fill the role of "intention" in the RAA. At this point, it is appropriate to ask: which of the survey questions Fishbein and Ajzen suggest researchers use to measure intention can be used to measure commonsense intention?

F&A write: "The readiness to act, represented by an intention, can find expression in such statements as the following:

- I will engage in the behavior.
- I intend to engage in the behavior.
- I expect to engage in the behavior.
- I plan to engage in the behavior.
- I will try to engage in the behavior." (F&A 2010, 39)

For the reasons discussed in section 3.2, statements of the first and third kind – those that begin with "I will"[30] and "I expect to" – being measures of behavioral expectation, are not good measures of commonsense intention. These measures must be rejected. Statements of the second kind,[31] which include a direct reference to the concept of intention, are clearly acceptable. Indeed, statements beginning "I intend…" are paradigm examples of statements of intention; any theory that advocated against their use as measures of intention would not be a theory of commonsense intention.

[30] For example, "I will exercise for at least 20 minutes, three times per week, for the next 3 months" (F&A 2010, 463).
[31] For example, "I intend to exercise for at least 20 minutes, three times per week, for the next 3 months" (ibid., 463).

Statements of the fourth kind, beginning "I plan…"[32], also seem to be acceptable. The statement "I intend to do it, but I do not plan to do it" seems to involve a contradiction, as does the converse ("I plan to do it, but I do not intend to do it"). It is important to note, however, that to plan to do something is not necessarily to *have a plan*, in the sense of "a detailed formulation of a program of action" (Merriam-Webster, 'plan'). It seems to me that statements beginning "I aim to…" are equally acceptable as measures of intention, as the statement "I intend to do it, but I do not aim to do it" also seems contradictory (as does, "I aim to do it, but I do not intend to do it").

Statements of the fifth kind – beginning "I will try…" – are a bit less straightforward. Trying does involve intentionality – if you're acting unintentionally, you're not trying. However, does saying that you will try always reflect a current intention? It seems not: in saying "I will try", I may be merely predicting a future intentional action which I do not currently intend to perform. For example, I might warn my friend, "I will try to convince you to give me the chocolate – if you want to help me stick to my resolution, you will need to ignore my future pleas." Does my saying this indicate that I currently intend to try to undo whatever chocolate-withholding arrangement I am presently making with my friend? Certainly not – it merely indicates that I know myself well enough to expect my intentions to change when my chocolate cravings kick in. (Similarly, statements regarding future intentions or future intentional actions do not imply present intentions. Consider the power-hungry king who, believing he will be kidnapped and brainwashed the following morning, says, "I do not intend to

[32] For example, "I plan to exercise for at least 20 minutes, three times per week, for the next 3 months" (ibid., 463).

give up the throne, but after the rebels brainwash me, I will have that intention", or "...I will intentionally do so.")

Elsewhere, F&A (2010, 463) also recommend the use of measures of intention of the form "I am willing to…" (e.g. " I am willing to exercise for at least 20 minutes, three times per week, for the next 3 months"). However, to be willing to do something is not to intend to do it. While a person may be willing to do either of two things she believes to be incompatible, it is not possible to intend to do two things you believe to be incompatible. (Here, where I say that a person is willing to do either of two incompatible things, X and Y, I mean that she is willing to do X and she is willing to do Y, rather than that she is willing to do "X or Y". For example, a person might say, "I'm willing to go to the library this evening, but I'm equally willing to stay home.") Thus, to be willing to do something is not to intend to do it. Measures of intention phrased in terms of willingness should not be used.

The above considerations indicate that if we wish to minimize noise and measure intentions as accurately as possible, only questions/statements that involve current intentions, plans, decisions, aims, and the like (e.g. "it is my goal to…"), and – perhaps – current trying (e.g. "I am trying to go to the gym three times a week") should be used. Thus, I propose we replace F&A's list with the following (partial) list of statements that express intention:

- I intend to engage in the behavior.
- I plan to engage in the behavior.
- I have decided to engage in the behavior.
- I aim to engage in the behavior.

3.5 Conclusion

I have argued that commonsense intentions are not behavioral expectations. If this is correct, "intentions" as defined within the RAA – i.e. intentions as behavioral expectations – are not commonsense intentions. Once we see that commonsense intentions are not behavioral expectations, it becomes clear that behavioral expectations are not well suited to the role played by "intentions" in the RAA. Thus, many of the measures that are commonly used by researchers working within the RAA framework are not appropriate. These should be replaced with measures that directly and uncontroversially assess intentions as we commonly think of them.

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, pp. 179–211.

Ajzen, I. (2012). Martin Fishbein's legacy: The reasoned action approach. *The Annals of the American Academy of Political and Social Science*, *640*, pp. 11–27.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, *2*(4), pp. 314–324.

Ajzen, I. & Fishbein, M. (1972). Attitudes and normative beliefs as factors influencing behavioral intentions. *Journal of Personality and Social Psychology*, *21*, pp. 1–9.

American Psychological Association. (n.d.). Informational influence. *APA Dictionary of Psychology*. Retrieved January 18, 2021, from *https://dictionary.apa.org/informational-influence*.

American Psychological Association. (n.d.). Normative influence. *APA Dictionary of Psychology*. Retrieved January 18, 2021, from *https://dictionary.apa.org/normative-influence*.

American Psychological Association. (n.d.). Normative. *APA Dictionary of Psychology*. Retrieved February 6, 2021, from *https://dictionary.apa.org/normative*.

American Psychological Association. (n.d.). Social influence. *APA Dictionary of Psychology*. Retrieved January 18, 2021, from *https://dictionary.apa.org/social-influence*.

American Psychological Association. (n.d.). Social pressure. *APA Dictionary of Psychology*. Retrieved February 5, 2021, from *https://dictionary.apa.org/social-pressure*.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: W.H. Freeman.

Biddle, B. J. (1979). *Role Theory: Expectations, Identities, and Behavior*. New York: Academic Press.

Biddle, B. J., & Thomas, E. J. (1966). *Role Theory: Concepts and research*. New York: John Wiley & Sons.

Booth, A. R., Norman, P., Harris, P. R., & Goyder, E. (2013). Using the theory of planned behaviour and self-identity to explain chlamydia testing intentions in young people living in deprived areas. *British Journal of Health Psychology*, *19*(1), pp. 101–112.

Bosnjak, M., Ajzen, I., & Schmidt, P. (2020). The theory of planned behavior: Selected recent advances and applications. *Europe's Journal of Psychology*, *16*(3), pp. 352–356.

Bratman, M. (1984). Two faces of intention. *The Philosophical Review*, *93*(3), pp. 375–405.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press. Reissued by CSLI Publications (1999).

Bratman, M. (1991). Cognitivism about practical reason (Review of *Practical Reflection*, by J. D. Velleman). *Ethics 102*(1), pp. 117–128.

Buchak, L. (2013). *Risk and Rationality*. New York: Oxford University Press.

Carfora, V., Caso, D., & Conner, M. (2016). The role of self-identity in predicting fruit and vegetable intake. *Appetite*, *106*, pp. 23–29.

Carfora, V., Cavallo, C., Caso, D., Del Giudice, T., De Devitiis, B., Viscecchia, R., Nardone, G., & Cicia, G. (2019). Explaining consumer purchase behavior for organic milk: Including trust and green self-identity within the theory of planned behavior. *Food Quality and Preference*, *76*, pp. 1–9.

Charng, H.-W., Piliavin, J. A., & Callero, P. L. (1988). Role identity and reasoned action in the prediction of repeated behavior. *Social Psychology Quarterly*, *51*(4), pp. 303–317.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, *28*(15), pp. 1429–1464.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, *28*(15), pp. 1429–1464.

Conner, M., & Norman, P. (2022). Understanding the intention-behavior gap: The role of intention strength. *Frontiers in Psychology*, 13.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), pp. 629–636.

Dulany, D. (1968). Awareness, rules, and propositional control: A confrontation with SR behavior theory. In T. Dixon & Deryck Horton (eds.), *Verbal Behavior and General Behavior Theory*, pp. 340–387. Englewood Cliffs, N.J.: Prentice-Hall.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fishbein, M., & Ajzen, I. (1981). On construct validity: A critique of Miniard and Cohen's paper. *Journal of Experimental Social Psychology*, *17*(3), pp. 340–350.

Fishbein, M., & Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York: Routledge.

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

French, J. R., & Raven, B. (1959). The bases of social power. In D. Cartwright, ed., *Studies in Social Power* (pp. 150-167). Ann Arbor: Research Center for Group Dynamics, Institute for Social Research, University of Michigan.

Gujarati, D. N. (2004). *Basic Econometrics*. 4th Edition, McGraw-Hill Companies.

Hale, J. L., Householder, B. J., & Greene, K. L. (2002). The theory of reasoned action. In J. P. Dillard & M. Pfau, eds., *The Persuasion Handbook: Developments in Theory and Practice*, pp. 259-286. Thousand Oaks, CA: SAGE Publications.

Harman, G. (1997). Practical reasoning. Reprinted in A. R. Mele, ed., *The Philosophy of Action* (Oxford: Oxford University Press), pp. 149–77.

Hassandra, M., Vlachopoulos, S. P., Kosmidou, E., Hatzigeorgiadis, A., Goudas, M., & Theodorakis, Y. (2011). Predicting students' intention to smoke by theory of planned behaviour variables and parental influences across school grade levels. *Psychology & Health*, *26*(9), pp. 1241–1258.

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2009). Cognitive burden of survey questions and Response Times: A psycholinguistic experiment. *Applied Cognitive Psychology*, *24*(7), pp. 1003–1020.

Mele, A. R. (1989). Intention, belief, and intentional action. *American Philosophical Quarterly*, *26*(1), pp. 19–30.

Mele, A. R. (2009). *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.

Mele, A. R. (2018). Free will, moral responsibility, and scientific epiphenomenalism. *Frontiers in Psychology*, 9.

Merriam-Webster. (n.d.). Intention. *Merriam-Webster.com dictionary*. Retrieved October 1, 2022, from *https://www.merriam-webster.com/dictionary/intention*.

Merriam-Webster. (n.d.). Norm. *Merriam-Webster.com dictionary*. Retrieved March 17, 2021, from *https://www.merriam-webster.com/dictionary/norm*.

Merriam-Webster. (n.d.). Normative. *Merriam-Webster.com dictionary*. Retrieved March 17, 2021, from *https://www.merriam-webster.com/dictionary/normative*.

Merriam-Webster. (n.d.). Plan. *Merriam-Webster.com dictionary*. Retrieved October 1, 2022 from *https://www.merriam-webster.com/dictionary/plan*.

Miniard, P. W., & Cohen, J. B. (1979). Isolating attitudinal and normative influences in behavioral intentions models. *Journal of Marketing Research*, *16*(1), pp. 102–110.

Miniard, P. W., & Cohen, J. B. (1981). An examination of the Fishbein-Ajzen behavioral intentions model's concepts and measures. *Journal of Experimental Social Psychology*, *17*(3), pp. 309–339.

O'Keefe, D. J. (2016). *Persuasion: Theory and Research*. (3rd ed.) Thousand Oaks, CA: SAGE Publications.

Park, H. S. (2000). Relationships among attitudes and subjective norms: Testing the theory of reasoned action across cultures. *Communication Studies*, *51*(2), pp. 162–175.

Raven, B. H. (1965). Social influence and power. In I. D. Steiner & M. Fishbein, eds., *Current Studies in Social Psychology* (pp. 371–382). New York: Holt, Rinehart, Winston.

Rise, J., Sheeran, P., & Hukkelberg, S. (2010). The role of self-identity in the theory of planned behavior: A meta-analysis. *Journal of Applied Social Psychology*, *40*(5), pp. 1085–1105.

Sabina del Castillo, E. J., Díaz Armas, R. J., & Gutiérrez Taño, D. (2021). An extended model of the theory of planned behaviour to predict local wine consumption intention and behaviour. *Foods*, *10*(9), 2187.

Setiya, K. (2008). Practical knowledge. *Ethics*, *118*(3), pp. 388–409.

Sheppard, B. H., et al. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research*, *15*(3), pp. 325–342.

Smetana, J. G., & Adler, N. E. (1980). Fishbein's value x expectancy model: An examination of some assumptions. *Personality and Social Psychology Bulletin*, *6*(1), pp. 89–96.

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2014). Time to retire the theory of planned behaviour. *Health Psychology Review*, *8*(1), pp. 1–7.

Sparks, P., & Guthrie, C. A. (1998). Self-identity and the theory of planned behavior: A useful addition or an unhelpful artifice? *Journal of Applied Social Psychology*, *28*(15), pp. 1393–1410.

Sparks, P., & Shepherd, R. (1992). Self-identity and the theory of planned behaviour: Assessing the role of identification with "green consumerism." *Social Psychology Quarterly*, *55*(4), pp. 388–399.

Steinmetz, H., Knappstein, M., Ajzen, I., Schmidt, P., & Kabst, R. (2016). How effective are behavior change interventions based on the theory of planned behavior? *Zeitschrift für Psychologie*, *224*(3), pp. 216–233.

Stryker, S., & Serpe, R. T. (1982). Commitment, identity salience, and role behavior: Theory and research example. In W. Ickes & E. S. Knowles, eds. *Personality, Roles, and Social Behavior*, pp. 199–218. New York: Springer.

Studer, B., & Knecht, S. (2016). A benefit–cost framework of motivation for a specific activity. In B. Studer & S. Knecht, eds., *Motivation: Theory, Neurobiology and Applications* (pp. 25–47). Amsterdam: Elsevier.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93.

Thoits, P. A., & Virshup, L. K. (1997). Me's and we's: Forms and functions of social identities. In R. D. Ashmore & L. J. Jussim, eds., *Self and Identity: Fundamental Issues* (pp. 106–133). New York: Oxford University Press.

Thorbjornsen, H., Pedersen, P. E., & Nysveen, H. (2007). "This is who I am": Identity expressiveness and the theory of planned behavior. *Psychology & Marketing*, *24*(9), 763–785.

Turner, J. C. (1991). *Social Influence*. Bristol, U.K.: Open University Press.

Velleman, J. D. (1985). Practical reflection. *The Philosophical Review*, *94*(1), pp. 33–61.

Velleman, J. D. (1989). *Practical Reflection*. Princeton, N.J.: Princeton University Press.

Wallace, R. J. (2001). Normativity, commitment, and instrumental reason. *Philosophers' Imprint*, *1*(3), pp. 1–26.

Warshaw, P. R., & Davis, F. D. (1985). Disentangling behavioral intention and behavioral expectation. *Journal of Experimental Social Psychology*, *21*(3), pp. 213–228.

Yaffe, G. (1995). Velleman on intentions as reasons for action. *Analysis*, *55*(2), pp. 107–115.

Zhu, J. (2003). Reclaiming volition: An alternative interpretation of Libet's experiment. *Journal of Consciousness Studies*, *10*(11), pp. 61–77.