"We Thinking" and its Consequences

By ROBERT AKERLOF[*]

Traditionally, economists have modeled agents as individualistic, uninfluenced by their social context, and motivated only by personal gain. However, increasingly, they have been drawing on concepts from outside economics – such as "norms," "esteem," and "identity" – to model agents' social natures.

A key reason for studying such social motivation is to shed light on the conditions that facilitate – or deter – collective action. It has been widely observed, for instance, that groups are more able to engage in collective action when they have a common, group identity (see Polletta and Jasper (2001) for a discussion). This paper will give one explanation for such a link. The paper will develop a new concept, "we thinking"; and it will also provide a deeper understanding of the concepts of norms, identity, and esteem.

I. "We thinking"

One way to solve collective action problems is by punishing noncooperation. Punishments serve to align individuals' incentives with those of the group. Economists have focused almost exclusively on solutions to collective action problems of this form.

I would define *true collective action* differently, however. When agents are truly engaged in collective action, they are *not acting individualistically*, but instead are motivated to pursue group ends.

I define "we thinking" as a mode of thinking in which an individual takes a group's goals as his own. Normally, agents are assumed to be individualistic; but a number of scholars have posited that agents can be induced to think in "we" – rather than in

"I" – terms. According to Searle (1990), for instance: "Collective intentional behavior is a primitive phenomenon which cannot be analyzed as just the summation of individual intentional behavior." But so far, models in which cooperation reflects true collective action have gained at most a small toehold in economic analysis (see Gold and Sugden (2007)).

There are countless real-world examples of situations in which individuals apparently act selflessly: in the best interest of their families, firms, teams, political parties, and countries. Social psychologists speak of "task cohesion," which refers to agents' ability to share a commitment to group goals. Military psychologists and sociologists have emphasized that soldiers who risk their lives typically are motivated by what is best for the group (see, for example, Shils and Janowitz (1948)).

It is clear that we-thinking can facilitate collective action, but what leads members of a group to think in we-terms rather than in I-terms? Various scholars (see, for instance, Tyler (1999)) have suggested that when there is a sense of *group pride*, agents are more likely to think in we-terms.

The explanation we shall offer here is that, when an agent thinks in we-terms, he is concerned with the esteem accorded the *group* rather than the esteem accorded himself as an individual. Therefore, the utility associated with we-thinking depends upon how the group is esteemed. Agents may not always have the freedom to choose whether to think in we-terms; most likely, we-thinking is cued rather than a choice. Nonetheless, if we-thinking is pleasurable, agents will seek out – rather than avoid – cues that induce we-thinking.

To summarize the argument so far, I have argued that group pride facilitates collective action because, when the group is a source of pride for agents, they are more

inclined to think in "we" terms.

Our objective, however, is to link identity to collective action. In this regard, this paper will further argue that group identity facilitates collective action because a common, group identity makes it possible for agents to take pride in the group. For instance, national pride seems to be based on a conception of a national identity. When a sports team wins a game, it is a source of pride to both team members and fans. I would argue that a sense of common identity underlies this ability to take pride in the team's accomplishments.

While we have sketched an argument as to why identity matters for collective action, a number of questions remain. What, precisely is identity? What makes it possible for a fan of a sports team to take pride in the team's accomplishments when he has no hand in bringing those accomplishments about? The remainder of this article will be concerned with: (1) defining identity and group pride/esteem more precisely, and (2) fleshing out why common, group identity is necessary for there to be group pride.

II. Norms

To understand identity, group pride, and the relationship between the two, we must first take a step back and discuss norms. How norms connect to pride and identity will become clear presently.

Norms are defined differently by different authors. A standard definition is that norms are judgments of what is better and worse behavior. But, one can also define norms more broadly as *judgments of any type*: judgments of actions, judgments of people, judgments of outcomes.

The types of norms that will be of interest to us are judgments of people and groups. The judgment of a group G can be represented by a real-valued function $N_i(s, G)$. This function answers the question: in the view of agent i, how good has group G been in state s?[1] In the event that a group G consists of a single individual ($G = \{j\}$),

$N_i(s, G)$ is, in fact, a judgment of an individual. I will use the notation $N_i(s, j)$ in place of $N_i(s, \{j\})$ to denote judgments of individuals.

The connection to esteem is relatively straightforward. $N_i(s, G)$ is an *absolute* judgment of how good an agent/group is. Esteem is a judgment of how good an agent/group is *relative to a comparison population* (P). The following is a mathematical representation of agent i's esteem for group G in state s ($E_i(s, G)$):

$$ E_i(s, G) = N_i(s, G) - \frac{1}{|P|} \sum_{G' \in P} N_i(s, G'). $$

This way of thinking about esteem is standard in social psychology (see R. Akerlof (2015) for a discussion). I would note that how the comparison group, P, is determined is an interesting question but it will not be a focus of this article.[2]

We argued in the previous section that, when an agent thinks in I-terms, he cares about how he is esteemed as an individual: that is, he cares about $E_i(s, i)$ (self-esteem) and perhaps $E_j(s, i)$ (the esteem he is accorded by other agents). When an agent thinks in we-terms – or, to be more precise, G-terms – he cares about how the group is esteemed: that is, he cares about $E_i(s, G)$ (the esteem he accords the group) and perhaps $E_j(s, G)$ (the esteem others accord the group). Agents will be inclined to think in we-terms when $E_i(s, G)$ and $E_j(s, G)$ are high (that is, when the group is a source of pride). Agents will also be inclined to think in we-terms when $E_i(s, i)$ and $E_j(s, i)$ are low (that is, when agents are accorded low individual esteem).

III. Identity

We have now linked norms to esteem. We still need to link norms to identity. Before we can do so, however, we must first define

---

[1] The state s should be thought of as inclusive of any variables that might be relevant for forming judgments.

[2] In Frank (1985)'s *Choosing the Right Pond*, agents compare themselves to those in their own pond. The pond is a choice in his model. In this sense, Frank (1985) is exploring aspects of how the comparison population P is determined.

identity.

I propose to define personal identity as "one's belief about one's type." Or, put another way, personal identity is *who one thinks one is*. Note that I use the term "type" to broadly refer to who a person is: it includes such features of a person as their character, ability, race, and gender.

I will denote an agent's type by θ_i. An agent's type is composed of traits (T): $\theta_i = \{\theta_i^t\}_{t \in T}$. Just as an individual has a type, we can think of a group G as having a type, consisting of those traits shared by group members: $\theta_G = \{\theta_i^t\}_{t \in T_G}$, where T_G denotes the shared traits of group members and $i \in G$.

We can define "group identity" in an analogous way to personal identity: as the belief of group members regarding the group's type. The term "identification" is also frequently used in the identity literature. We can define "identification" as a belief that one shares traits with other agents.

I would note that, in G. Akerlof and Kranton (2000)'s model of identity, one's identity may be a choice. Their way of thinking about identity is not wholly inconsistent since it may be possible to choose what one believes about one's type.[3]

In the next section of the paper, I will argue that how agents are judged/esteemed will depend upon their types. Consequently, a link exists between norms/esteem, on the one hand, and identity on the other hand.

IV. Judgments and Type

I would now like to examine more carefully how agents form judgments of people/groups. It is useful to think of agents as applying a set of principles in forming judgments: $N_i(s, G) = \sum_{D \in \mathcal{D}_i^G} D(s, G)$. $D \in \mathcal{D}_i^G$ denotes a principle agent i uses in judging group G; $\mathcal{D}_i^G \subseteq \mathcal{D}_i$ denotes the set of principles, among all those used by agent i ($\mathcal{D}_i$), that apply to group G.

It is clear that agents do not use identical principles in forming judgments. For instance, one agent might particularly weight athletic achievement while another agent might particularly weight academic achievement. Despite this lack of uniformity, I claim that agents apply principles that they believe have a certain property (Property P).

DEFINITION 1: *A principle has Property P if it always judges an agent/group of a given type the same way.*[4,5]

To build some intuition, consider an example. Suppose the performance of athletes (p_i) is given by: $p_i = a_i \cdot \theta_i \cdot w_i$, where a_i is an athlete's action choice, θ_i is an athlete's type, and w_i is the weather. Consider the principle $D(s, j) = w_j$ which judges athletes solely on the weather. If this principle were applied, an athlete of a given type (θ_j) would be judged differently depending upon whether it happens to be rainy or sunny. Therefore, Property P is violated.

A principle that clearly possesses Property P is one that judges athletes based

---

[3]G. Akerlof and Kranton (2000) suggest that tied to identities are "prescriptions" regarding appropriate behavior. For example, many people view it as appropriate for girls to play with dolls but inappropriate for boys to do so. The framework proposed here can also capture the idea of prescriptions. I would suggest that agents hold beliefs regarding what constitutes "natural behavior" for a person of a given type. It is a norm that agents should behave naturally (i.e., agents are judged to be worse, all else equal, when their behavior is less natural). Returning to the doll example, the idea is that it would be viewed as appropriate for girls to play with dolls because it is *natural* for them to do so; it would be viewed as inappropriate for boys to play with dolls because it is *unnatural* for them to do so.

[4]In ethics, there is a related literature on "moral luck." Suppose, for instance, drunk driver A is lucky and does not have an accident while drunk driver B is unlucky and does. Ethicists have observed that A and B might be judged differently even if they are of the same type ($\theta_A = \theta_B$). They have raised the issue of moral luck precisely because they see it as an exception to the general rule that agents of a given type are judged the same. Note that I am claiming that agents *believe* their principles have Property P – not that they always do. This allows for some violations as described by ethicists.

[5]There are a number of stories – to give one example, *The Prince and the Pauper* – in which characters trade places or are switched at birth. These stories criticize an existing social order by demonstrating violations of Property P (i.e., the prince and pauper are of similar type but are judged differently).

upon type alone: $D(s, j) = \theta_j$. However, this is not necessarily the only such principle. Suppose, for instance, all athletes choose $a_j = 1$. Then, $\frac{p_j}{w_j} = \theta_j$. Therefore, $D(s, j) = \frac{p_j}{w_j}$ also has Property P.

An interesting feature of this principle is that, *ex post*, an athlete is judged on his type, but he is *not* judged on his type *ex ante*. *Ex ante*, the athlete is judged on his performance (discounting the effect of weather); moreover, the athlete's choice of action affects how he is judged.

In the existing economics literature, I would note that a number of papers have assumed agents are judged based upon type alone: for instance, Bénabou and Tirole (2002). However, other papers – for instance, Bénabou and Tirole (2011) – assume that there are better and worse actions and that agents are judged to be better (worse) when the actions they take are better (worse). Property P says that both types of judgment are admissible; however, when people are judged *ex ante* on their actions, they must still be judged *ex post* on type.

Property P requires that a principle must judge two agents/groups, G and G', the same if they are of the same type. However, it does not follow from Property P that the overall judgments of two agents/groups ($N_i(s, G)$ and $N_i(s, G')$) must be the same since different principles might be applied to G and G'. Property Q, defined below, is an elaboration of Property P that ensures overall judgments of people/groups of the same type will be the same. I claim that agents, in fact, apply principles that they believe meet Property Q.

DEFINITION 2: *A principle D has Property Q if there exist traits, T_D, such that: (1) D applies to group G if and only if its members share traits T_D; and (2) Two groups, G and G', that share traits T_D are always judged the same.*

Property Q explains our basic puzzle of why common, group identity is necessary for there to be group pride. An example will help to illustrate. When the Yankees win the World Series (as they often do), it is a source of pride to many New Yorkers. What underlies such pride-taking? It seems clear that New Yorkers are applying a principle, $D_{baseball}$, that judges cities on the performance of their sports teams. New Yorkers take pride in their city because $D_{baseball}$ judges their city well relative to other cities.

Property Q says that, in order to apply the principle $D_{baseball}$ to cities, New Yorkers must believe that: (1) people within cities share certain traits ($T_{baseball}$), and (2) how a city's baseball team performs is a function of these shared traits. (1) is the key point to be emphasized. Framed differently, (1) says that, to take pride in New York, New Yorkers must believe their city and comparison cities have distinct, group identities.

Property Q has another implication of interest. It says that a principle D that applies to a group G has the property that $D(s, G) = D(s, i)$ for all $i \in G$. In other words, principles that judge groups also judge group members. Applied to our example, this means that if New York (as a whole) is judged/esteemed on how the Yankees perform, so too are New Yorkers individually. Because of Property Q, we can think of an individual's esteem as deriving from various components of his identity. A particular New Yorker will derive some esteem from his identity as a New Yorker (i.e., from principles that apply to the group of New Yorkers, to which he belongs); he will also, most likely, derive esteem from other aspects of his identity (his family, his workplace, and also aspects of his identity that are idiosyncratic to himself). Returning to our discussion of we-thinking, the implication is that the willingness to engage in G-thinking will depend upon how much esteem an agent derives from his identity as part of group G relative to other parts of his identity.

V. Implications for Identity Formation

We have accomplished the main aim of this article: namely, to show why groups with a sense of common identity find it easier to engage in collective action. It is worth noting, though, that the framework may be useful for understanding how agents' prin-

ciples and identities form. To the extent that identities and principles ($\mathcal{D}_i$) are chosen, the framework suggests that esteem considerations – the desire for self-esteem and peer esteem – may play an important role in shaping them.

R. Akerlof (2015) presents a model in which agents choose their principles. A tension exists in the model between agents' desire, on the one hand, to adopt the same principles as peers (i.e., conform) for the sake of peer esteem and a desire to adopt different principles (i.e., differentiate) for the sake of self-esteem.

I would suggest that a similar tension exists in the choice of identities. There is a desire, on the one hand, to differentiate in one's choice of identity. If one believes oneself to be of exactly the same type as everyone else, Property Q requires that one judge oneself the same. In order to have positive self-esteem ($E_i(s, i) > 0$), it is necessary to believe that people are, at least in some ways, different. New Yorkers must view themselves as different from Bostonians, for example, in order to think of themselves as superior. On the other hand, there are incentives for agents to form identity groups. Agents are normally inclined to disesteem one another since only by esteeming others less can one raise one's self-esteem; but, when agents identify with one another, it creates incentives to positively esteem one another: there is a mutual desire to judge the shared type well. New Yorkers enjoy, for example, being able to engage in mutual pride-taking over a win by the Yankees. Therefore, just as agents' desire for self-esteem motivates disidentification with Boston, the desire for peer esteem motivates New Yorkers to forge a common, group identity.[6]

Agents' tendency to sort into what psy-

chologists call "ingroups" and "outgroups" has important implications for collective action. It means that, almost inevitably, some groups will find it easy to engage in collective action while others will find it difficult.

REFERENCES

Akerlof, George, and Rachel Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics*, 715–753.

Akerlof, Robert. 2015. "Value Formation: The Role of Esteem." *Working Paper.*

Bénabou, Roland, and Jean Tirole. 2002. "Self-confidence and Personal Motivation." *Quarterly Journal of Economics*, 871–915.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics*, 126(2): 805–855.

Frank, Robert. 1985. *Choosing The Right Pond: Human Behavior and The Quest for Status.* Oxford University Press.

Gold, Natalie, and Robert Sugden. 2007. "Collective intentions and team agency." *The Journal of Philosophy*, 109–137.

Polletta, Francesca, and James M Jasper. 2001. "Collective Identity and Social Movements." *Annual Review of Sociology*, 283–305.

Searle, John. 1990. "Collective Intentions and Actions." In *Intentions in Communication.* , ed. P. Cohen, J. Morgan and M. E. Pollack. Cambridge, MA:Bradford Books, MIT Press.

Shils, Edward, and Morris Janowitz. 1948. "Cohesion and Disintegration in the Wehrmacht in World War II." *Public Opinion Quarterly*, 12(2): 280–315.

Tyler, Tom. 1999. "Why People Cooperate with Organizations: An Identity-Based Perspective." In *Research in Organizational Behavior Vol. 21.* , ed. B M Staw and R Sutton, 201–246. Greenwich, CT:JAI Press.

---

[6]In a series of influential experiments, Henri Tajfel and coauthors have found that even randomly assigning subjects to groups leads subjects to identify with those in their own group and view those in their own group as superior (i.e., worthy of greater esteem). This finding aligns with the perspective I have just given in which identities are motivated by esteem considerations. I would suggest that the experimental assignment has an effect because it coordinates agents to engage in mutual pride-taking along certain lines.